



HAL
open science

BRIEF-based mid-level representations for time series classification

Renato Souza, Raquel Almeida, Roberto Miranda, Zenilton Kleber Gonçalves Do Patrocinio, Simon Malinowski, Silvio Jamil F. Guimarães

► **To cite this version:**

Renato Souza, Raquel Almeida, Roberto Miranda, Zenilton Kleber Gonçalves Do Patrocinio, Simon Malinowski, et al.. BRIEF-based mid-level representations for time series classification. CIARP 2019 - 24th Iberoamerican Congress on Pattern Recognition, Oct 2019, La Havane, Cuba. pp.449-457, 10.1007/978-3-030-33904-3_42 . hal-02371260

HAL Id: hal-02371260

<https://inria.hal.science/hal-02371260>

Submitted on 19 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BRIEF-based mid-level representations for time series classification

Renato Souza¹, Raquel Almeida¹, Roberto Miranda¹, Zenilton Kleber G. do Patrocinio Jr¹, Simon Malinowski², and Silvio Jamil F. Guimarães¹

¹ Pontifical Catholic University of Minas Gerais, Computer Science Department
Belo Horizonte-MG, Brazil

² Université de Rennes 1, IRISA, Rennes, France

Abstract. Time series classification has been widely explored over the last years. Amongst the best approaches for that task, many are based on the Bag-of-Words framework, in which time series are transformed into a histogram of word occurrences. These words represent quantized features that are extracted beforehand. In this paper, we aim to evaluate the use of accurate mid-level representation called BossaNova in order to enhance the Bag-of-Words representation and to propose a new binary time series descriptor, called BRIEF-based descriptor. More precisely, this kind of representation enables to reduce the loss induced by feature quantization. Experiments show that this representation in conjunction to BRIEF-based descriptor is statistically equivalent to traditional Bag-of-Words, in terms time series classification accuracy, being about 4 times faster. Furthermore, it is very competitive when compared to the state-of-the-art.

Keywords: Time series · Mid-level representations · BRIEF-based descriptors

1 Introduction

Time series can be seen as series of ordered measurements. They contain temporal information that needs to be taken into account when dealing with such data. Time series classification (TSC) could be defined as follows: given a collection of unlabeled time series, one should assign each time series to one of a predefined set of classes. TSC is a challenge that is receiving more and more attention recently due to its most diverse applications in real life problems involving, for example, data mining, statistics, machine learning and image processing.

An extensive comparison of TSC approaches is performed in [3] and an evaluation of mid-level representations in TSC is given by [1]. Three particular methods stand out from other core classifiers for their accuracies: COTE [4], BOSS [13] and D-VLAD [1]. According to [4], COTE contains classifiers constructed in the time, frequency, change, and shapelet transformation domains combined in alternative ensemble structures. BOSS is a dictionary-based approach that adopts an extraction of Fourier coefficients from time

series windows, and D-VLAD is also a dictionary-based in which SIFT-based descriptors are assembled by the mid-level presentation called VLAD [11]. Many other dictionary-based approaches have been proposed and used recently [1,5,6]. These methods share the same overall steps: (i) extraction of feature vectors from time series; (ii) creation of a codebook (composed of codewords) from extracted feature vectors; and (iii) representation of time series using extracted codewords.

The first and third steps are very important to design an accurate TSC scheme. In this paper, we propose a dictionary-based approach for TSC, that follows these steps. For the feature vector extraction, we propose to adapt a binary descriptor that was designed for image description, which is called BRIEF [9]. This descriptor has the main advantage to be fast to compute, while providing an accurate description. For the representation step, it has been shown in [1], that the classical Bag-of-Word representation (in which codewords are quantized) could be enhanced by using more discriminative methods. In this paper, we make use of BossaNova [2], which is a method that keeps more information than a traditional Bag-of-Word approach.

The main contributions of this paper are hence two-folds: (i) adaptation of a binary descriptor for describing time series, called BRIEF, and (ii) the use of BossaNova in order to enrich the final time series representation.

This paper is organized as follows. Section 2 describes some related works about time series classification. In Section 3, we present a methodology for time series classification by using powerful mid-level representation built on BRIEF-based descriptors. Section 4 details the experimental setup and results to validate the method, and finally, some conclusions are drawn in Section 5.

2 Related work

In this section, we give an overview about the related work on TSC. One of the earliest methods for that task is the combination of 1-nearest-neighbor classifier with Dynamic Time Warping. It has been a baseline for TSC for many years thanks to its good performance. Recently, more sophisticated approaches have been designed for TSC.

Shapelets, for instance, were introduced in [14]. They represent existing subsequences able to discriminate classes. Hills *et al.* proposed the shapelet transform [10], which consists in transforming a time series into a vector whose components represent the distance between time series and different shapelets, extracted beforehand. Classifiers, such as SVM, can then be used with these vectorial representations of time series.

Numerous approaches have been designed based on the Bag-of-Word (BoW) framework. This framework consists in extracting feature vectors from time series, creating a dictionary of words using these extracted features, and then representing each time series as a histogram of words occurrence. The different approaches proposed in the literature differ mainly on the kind of features that are extracted. Local features such as mean, variance and extrema are considered in [6], Fourier coefficients in [13]; while SAX coefficients are used in [12]. Recently,

SIFT-based descriptors adapted to time series have been considered as feature vectors in [5].

All the methods based on the BoW framework create a dictionary of words by quantizing the set (or a subset) of extracted features. This quantization step induces some loss when representing time series as a histogram of words occurrence. In [1], in order to improve the accuracy of time series representations, the authors have studied several more discriminative mid-level representations, such as VLAD for instance.

3 Time series classification based on BRIEF descriptor

In this section, we describe the proposed TSC scheme. We aim at improving classical BoW representation for time series in two ways: we make use of a binary descriptor adapted from BRIEF, that is very fast to compute; and we use BossaNova in order to enrich the final representation of time series. The use of BRIEF is motivated by the study made in [8] in which a comparison of different low-level descriptors to classify pornography videos presented competitive results taking much less time.

Due to the low complexity, binary descriptors is mostly used in real-time applications due to the simplicity of computational procedure not only for the descriptor itself but also for its similarity measure. The basic idea of binary descriptors is to encode some information of a path into a binary sequence, by comparing the intensity of the points present in that path. In the case of BRIEF, there is neither sampling pattern nor orientation compensation.

The proposed approach is composed of the following steps: (i) keypoints selection; (ii) keypoints description; (iii) generation of final mid-level representation of time series; and (iv) classification. These steps are detailed below, and first two are illustrated in Fig. 1. In the following, let $X = x_1, \dots, x_n$ be a time series of length n .

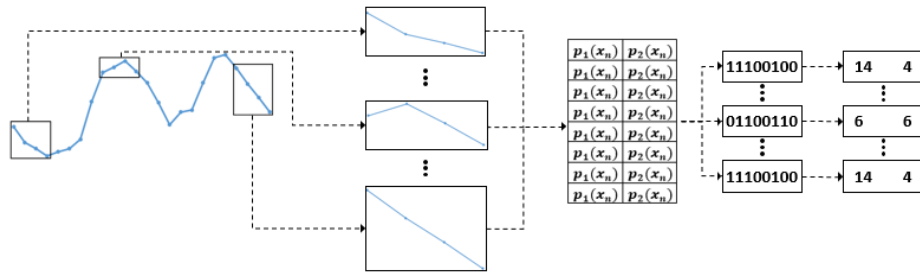


Fig. 1. Firstly, using a time series S , its created windows according to the size s and the distance between the key points. With the windows and selected n pairs randomly. The binary sequence constructed are concatenate into bit-string. Then, the binary strings are converted to a integer-string where a integer is computed for each k bits.

Keypoints selection: We start by selecting the keypoints that will be described at the next step. Dense selection of keypoints have shown to be more efficient than other methods. We hence select keypoints regularly inside the time series: one keypoint is selected every time interval of τ instants, in which τ is a parameter of the method. At the end of this step, the set $\{x_1, x_{1+\tau}, \dots\}$ of keypoints is selected.

Keypoints description: Let x_k be the keypoint that we want to describe. For that purpose, the window $W = s_1, \dots, s_w$ of length w is selected around x_k . Then p pairs of numbers $(i_1, i_2) \in [1, w]^2, i_1 \neq i_2$, are then randomly selected. Note that the same pairs are kept to describe each keypoints. For each pair (i_1, i_2) , a binary number $b_{(i_1, i_2)}$ is computed as follows :

$$b_{(i_1, i_2)} = \begin{cases} 1, & \text{if } s_{i_1} < s_{i_2}; \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

When all pairs have been processed, a binary vector of length p is generated and represents the description (feature vector) of the keypoint x_k .

Mid-level representation: Let $\mathbb{X} = \{\mathbf{x}_j \in \mathbb{R}^d\}_{j=1}^N$ be an unordered set of d -dimensional descriptors \mathbf{x}_j extracted from the data. Let also $\mathbb{C} = \{\mathbf{c}_m \in \mathbb{R}^d\}_{m=1}^M$ be the codebook learned by an unsupervised clustering algorithm, composed by a set of M codewords, also called prototypes or representatives. Consider $\mathbb{Z} \in \mathbb{R}^M$ as the final vector mid-level representation. As formalized in [7], the mapping from \mathbb{X} to \mathbb{Z} can be decomposed into three successive steps: (i) coding; (ii) pooling; and (iii) concatenation. In order to keep more information than BoW during pooling step, we have used BossaNova [2] as mid-level representation which follows BoW formalism (coding/pooling). It uses a density-based pooling strategy and a localized soft-assignment coding that considers only the k -nearest codewords for coding a local descriptor. To keep more information than the BoW during the pooling step, BossaNova pooling function, g , estimates the probability density function of α_m : $g(\alpha_m) = \text{pdf}(\alpha_m)$, by computing the following histogram of distances $z_{m,b}$:

$$z_{m,b} = \text{card} \left(\mathbf{x}_j \mid \alpha_{m,j} \in \left[\frac{b}{B}; \frac{b+1}{B} \right] \right), \quad (2)$$

$$\frac{b}{B} \geq \alpha_m^{\min} \quad \text{and} \quad \frac{b+1}{B} \leq \alpha_m^{\max}, \quad (3)$$

in which:

- B denotes the number of bins of each histogram z_m ;
- $\alpha_{m,j}$ represents a dissimilarity measure between codewords and feature points x_j ; and
- $\alpha_m^{\min}, \alpha_m^{\max}$ limits the range of distances for the descriptors considered in the histogram computation.

The final BossaNova representation is in the form:

$$\mathbf{z} = [[z_{m,b}], t_m]^T, \quad (m, b) \in \{1, \dots, M\} \times \{1, \dots, B\}, \quad (4)$$

in which t_m scalar value for each codeword as an approximation of the traditional BoW representation.

Classification: The mid-level representation for each time series is then passed to a classifier to learn how to discriminate classes using this description.

4 Experimental analysis

In this section, we describe our experiments in order to investigate the impact, in terms of classification performances, of more powerful encoding methods applied to dense extracted features for TSC.

4.1 Experimental setup

Experiments are conducted on the 84 currently available datasets from the UCR repository, the largest on-line database for time series classification. Due to problems on feature extraction, we ignored the 12 largest datasets. All datasets are splitted into training and test sets, whose sizes vary between less than 20 and more than 8,000 time series. For a given dataset, all time series have the same length, ranging from 24 to more than 2,500 points. In order to compute the mid-level representation, we have extracted SIFT-based descriptors as proposed in [5] and BRIEF-based descriptors by using dense sampling. Codebooks have been generated using the following number of clusters $\{16, 32, 64, 128, 256, 512\}$. The sizes of the window and the numbers of pairs of the BRIEF-based descriptor are selected from $\{16, 32, 64, 128, 256, 512, 1024\}$, respectively. The representations are normalized by a L2-norm and signed square root. Moreover, we have used d-BRIEF and d-SIFT for indicating the step used in the keypoint extraction (sampling rate); for example, 1-BRIEF means that all points of the time series are extracted. The best sets of parameters are obtained by a 5-fold cross-validation to be used with a RBF kernel SVM model during classification step.

4.2 Quantitative analysis

Despite the interest in classifying time series with high accuracy, it is very important to propose methods with low computational time. In that sense, feature extraction for TSC using BRIEF-based descriptor presents very competitive results when compared to SIFT-based descriptor. Depending on the sampling rate, the speed-up is 3.64 and 5.91 times, for 1-BRIEF and 2-BRIEF, respectively, as can be seen in Table 1.

In Fig. 2, we present the accuracy results for compared methods. As we can see, the 1-VLAD (SIFT-based dense sampling and VLAD representation) when compared to COTE presents very competitive results (when COTE and

1-VLAD are computed by using the same experiment protocol). Regarding 1-SIFT, the results are competitive to BOSS. When 1-BRIEF is compared to BoTSW [5] and 1-SIFT, the proposed binary descriptor to time series have presented very competitive results in terms of accuracy (as illustrated in Fig. 2) but it is about 4 times faster for keypoint description. Furthermore, they are statistically equivalent when the results are compared by using a t-student test with 95% of confidence level.

We also have compared the performance of the methods by using a pair-wise distribution, as illustrated in Fig. 3. Taking into account these distribution of points, it is possible to argue that the BRIEF-based and SIFT-based descriptors present similar results.

4.3 Comparison to the state-of-the-art methods

In order to study the impact of specially designed mid-level representations on TSC. We focus on two different analysis. In the first one, we compare the studied representations, namely d-BRIEF and 1-SIFT to BoTSW and 1-VLAD, which are our baselines. In the second one, we present a comparative analysis between

Time (in ms)	1-BRIEF	2-BRIEF	1-SIFT
Total	1,012,776.00	623,836.00	3,691,278.00
Average per dataset	14,677.91	9,041.10	53,496.78
Speed-up	3.64	5.91	—

Table 1. Average time for describing the datasets using BRIEF-based and SIFT-based descriptors. Here, 2-BRIEF means that a BRIEF descriptor is computed for each 2 keypoints and 1-BRIEF (1-SIFT) for all points of the time series.

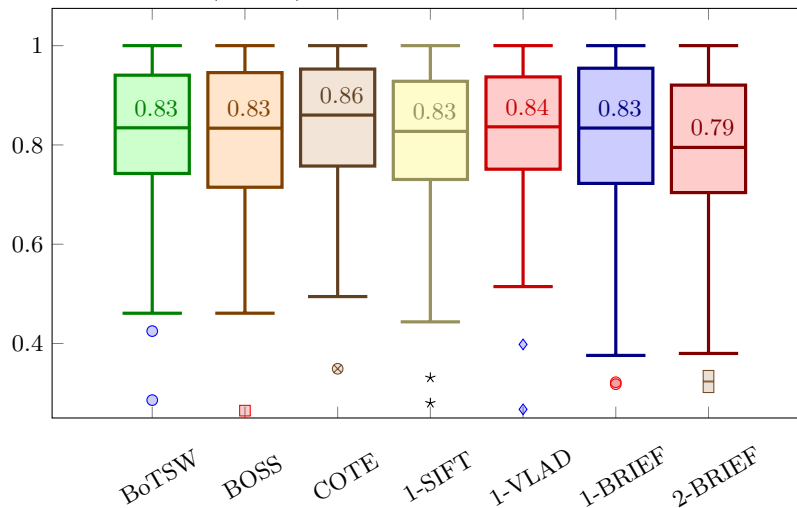


Fig. 2. Accuracy comparison of different methods applied on TSC.

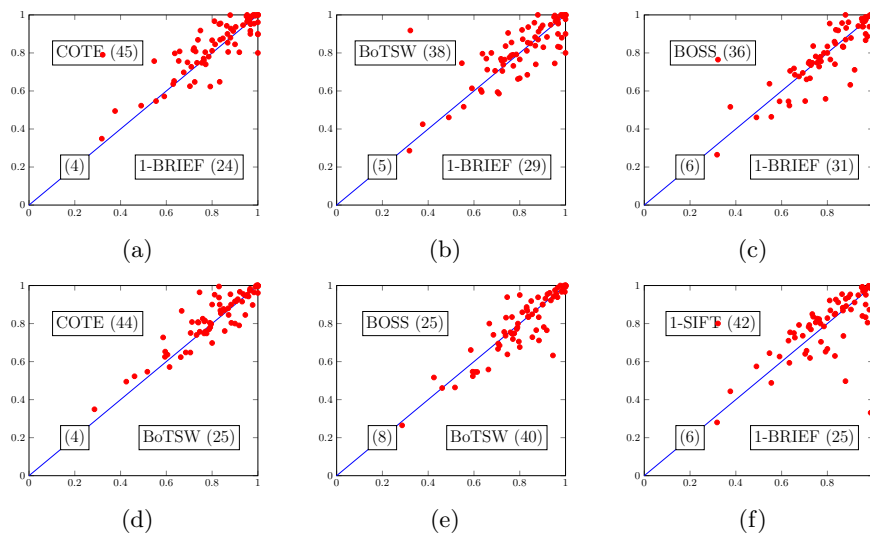


Fig. 3. Pairwise comparison of classification rates between BRIEF-based mid-level representations and the state-of-the-art.

the state-of-the-art, namely BoTSW [5], BOSS [13] and COTE [4], and the new proposed use of mid-level description.

In both cases, we used the average accuracy rate and rank that are summarized in Table 2. As illustrated, 1-VLAD obtained the best results among the mid-representation, and it is very competitive to the state-of-the-art, being better than BoTSW and BOSS. When compared to our baselines BoTSW and 1-VLAD, the 1-BRIEF is statistically equivalent taking into account the paired t-test with 95% of confidence level. 1-BRIEF is statistically better than 2-BRIEF, and equivalent to 1-SIFT. Furthermore, concerning the comparison of 1-BRIEF to BoTSW and 1-SIFT, we have observed: (i) the binary descriptor in conjunction to BossaNova presented competitive performances but it is faster, which confirms our initial assumptions; and (ii) 1-BRIEF presented good results in terms of classification rates and average rank but it is worse than 1-SIFT and better than BOSS.

Method	State-of-the-art				Mid-level representations		
	BoTSW	BOSS	COTE	1-VLAD	1-SIFT	2-BRIEF	1-BRIEF
Rate	0.819	0.809	0.836	0.827	0.807	0.780	0.813
Ranking	3.78	4.21	3.04	3.27	3.55	4.71	4.01

Table 2. Comparison to the state-of-the-art in terms of classification rates and ranking.

5 Conclusions

Time series classification is a challenge task due to its most diverse applications in real life. Among the several kind of approaches, dictionary-based ones have received much attention in the last years. In a general way, these ones are based on the extraction of feature vectors from time series, creation of codebook from the extracted feature vectors and finally representation of time series as traditional Bag-of-Words. In this work, we studied the impact of more discriminative and accurate mid-level representation, called BossaNova, for describing the time series taking into account SIFT-based descriptors [5] and the proposed binary descriptor for time series, called BRIEF-based descriptor.

According to our experiments, 1-BRIEF is statistically equivalent to 1-SIFT and BoTSW but the binary keypoint description is about 4 times faster than the non-binary one. Moreover, we achieve competitive results when compared to the some state-of-the-art methods, mainly with BOSS and 1-VLAD. However, despite the pairwise comparison (Fig. 3) involving both methods, COTE is slightly better than 1-VLAD in terms of average rank but both are statistically equivalent when the comparison is done by the paired t-test with 95% of confidence. Thus, the use of more accurate mid-level representation in conjunction with BRIEF-based descriptor seems to be a very interesting approach to cope with time series classification. From our results and observations, we believe that a future study of the normalization and distance functions could be interesting in order to understand their impact in our method, since according to [11] the reduction of frequent codeword influence could be profitable.

Acknowledgments

This study was financed in part by the Coordenao de Aperfeioamento de Pessoal de Nvel Superior - Brasil (CAPES) - Finance Code 001. Moreover, the authors are grateful to PUC Minas, FAPEMIG and the TRANSFORM project funded by CAPES/STIC-AMSUD (18-STIC-09) for the partial financial support to this work.

References

1. R. Almeida, H. Herlanin, Z. K. G. do Patrocínio Jr., S. Malinowski, and S. J. F. Guimarães. Evaluation of bag-of-word performance for time series classification using discriminative sift-based mid-level representations. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 23rd Iberoamerican Congress, CIARP 2018, Madrid, Spain, November 19-22, 2018, Proceedings*, pages 109–116, 2018.
2. S. Avila, N. Thome, M. Cord, E. Valle, and A. d. A. Araújo. Pooling in image representation: The visual codeword point of view. *Computer Vision and Image Understanding*, 117(5):453–465, 2013.

3. 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*, 31(3):606–660, May 2017.
4. A. Bagnall, J. Lines, J. Hills, and A. Bostrom. Time-series classification with cote: The collective of transformation-based ensembles. *IEEE Transactions on Knowledge and Data Engineering*, 27(9):2522–2535, Sept 2015.
5. A. Bailly, S. Malinowski, R. Tavenard, L. Chapel, and T. Guyet. *Dense Bag-of-Temporal-SIFT-Words for Time Series Classification*, pages 17–30. Springer International Publishing, Cham, 2016.
6. M. G. Baydogan, G. Runger, and E. Tuv. A bag-of-features framework to classify time series. *IEEE PAMI*, 35(11):2796–2802, 2013.
7. Y. L. Boureau, F. Bach, Y. LeCun, and J. Ponce. Learning mid-level features for recognition. In *Proceedings of the CVPR 2010*, pages 2559–2566, June 2010.
8. C. Caetano, S. Avila, S. Guimaraes, and A. d. A. Araújo. Pornography detection using bossanova video descriptor. In *Proceedings of the EUSIPCO 2014*, pages 1681–1685, Lisbon, Portugal, 2014. IEEE.
9. M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. In *European conference on computer vision*, pages 778–792. Springer, 2010.
10. J. Hills, J. Lines, E. Baranauskas, J. Mapp, and A. Bagnall. Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery*, 28(4):851–881, 2014.
11. H. Jegou, F. Perronnin, M. Douze, J. Sánchez, P. Perez, and C. Schmid. Aggregating local image descriptors into compact codes. *IEEE PAMI*, 34(9):1704–1716, 2012.
12. J. Lin, R. Khade, and Y. Li. Rotation-invariant similarity in time series using bag-of-patterns representation. *Journal of Intelligent Information Systems*, 39(2):287–315, 2012.
13. P. Schäfer. The BOSS is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, 29(6):1505–1530, 2014.
14. L. Ye and E. Keogh. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 947–956. ACM, 2009.