Biomedical Signal Processing and Control 8 (2013) 634–644

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Bag-of-words representation for biomedical time series classification

Jin Wang[a,c,*], Ping Liu[b], Mary F.H. She[a,c], Saeid Nahavandi[a], Abbas Kouzani[d]

[a] Center for Intelligent Systems Research, Deakin University, Waurn Ponds 3217, Australia
[b] Department of Computer Science, University of South Carolina, Columbia, SC 29205, USA
[c] Institute for Frontier Materials, Deakin University, Waurn Ponds 3217, Australia
[d] School of Engineering, Deakin University, Waurn Ponds 3217, Australia

ARTICLE INFO

Article history:
Received 30 January 2013
Received in revised form 14 June 2013
Accepted 14 June 2013
Available online 12 July 2013

Keywords:
Bag of words
Codebook construction
k-Means clustering
EEG
ECG

ABSTRACT

Automatic analysis of biomedical time series such as electroencephalogram (EEG) and electrocardiographic (ECG) signals has attracted great interest in the community of biomedical engineering due to its important applications in medicine. In this work, a simple yet effective bag-of-words representation that is originally developed for text document analysis is extended for biomedical time series representation. In particular, similar to the bag-of-words model used in text document domain, the proposed method treats a time series as a text document and extracts local segments from the time series as words. The biomedical time series is then represented as a histogram of codewords, each entry of which is the count of a codeword appeared in the time series. Although the temporal order of the local segments is ignored, the bag-of-words representation is able to capture high-level structural information because both local and global structural information are well utilized. The performance of the bag-of-words model is validated on three datasets extracted from real EEG and ECG signals. The experimental results demonstrate that the proposed method is not only insensitive to parameters of the bag-of-words model such as local segment length and codebook size, but also robust to noise.

1. Introduction

With the development of modern technology and reduction of hardware cost, a large amount of biomedical signals such as electroencephalogram (EEG) and electrocardiographic (ECG) are collected every day. These biomedical signals are very useful for monitoring human's physical condition. It is however a challenging task to efficiently and effectively analyze these signals. Traditionally, these signals are manually analyzed by professional experts. However, there are several disadvantages of the manual analysis. Firstly, comparing to the large amount of biomedical signals, the number of professional experts, especially the ones with extensive experience is very limited. Secondly, inspection and monitoring of long-term biomedical signals such as EEG and ECG signals are always very time consuming. It is difficult to keep a high level of concentration during a lengthy inspection, giving rise to an increase in the false hit rate by the operator. Finally, it is frequently needed to find inter-reader variability in the manual inspection and monitoring by experts. Therefore, an automated system that can assist professional experts to analyze long-term biomedical signals is very valuable in real-word applications.

Automatic analysis of biomedical time series such as EEG and ECG signals based on machine learning techniques has been applied to a variety of real-word applications. For instance, EEG signals are automatically analyzed for epileptic seizure detection [1–3], brain computer interaction [4–6], human mental fatigue detection [7] and emotion recognition [8]. ECG signals that provide useful information about heart rhythm are used to study heart arrhythmias [9,10]. It is essential to extract meaningful features to represent individual time series in the aforementioned applications. Some methods [11,12] directly describe time series in time domain while some others extract features from transformed domain [13,14,10]. For instance, Zadeh et al. [12] extracted morphological and timing-interval features from ECG segments to classify heartbeats. Guo et al. [13] extracted line length features based on Discrete Wavelet Transform (DWT) to detect epileptic EEG segments.

Most of the previous representations extract local temporal or frequency information to characterize time series, which are very effective for short time series or time series with periodic waveforms. However, they may have limited ability to capture structural similarity of long time series which have repetitive waveforms, for instance, electrocardiography (ECG) and electroencephalography (EEG) signals. In order to capture the high-level structural information of time series, Lin [15] proposed a bag-of-patterns (BoP) representation by converting a time series to a words string using the Symbolic Aggregate approXimation (SAX). The temporal order of local segments, i.e., local patterns, in a time series is ignored

* Corresponding author at: Center for Intelligent Systems Research, Deakin University, Waurn Ponds 3217, Australia. Tel.: +61 406417698.
E-mail addresses: jay.wangjin@gmail.com, wjin@deakin.edu.au (J. Wang).

1746-8094/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2013.06.004

and all the local segments in the time series are histogrammed to construct a bag-of-patterns representation. The bag-of-patterns representation is effective to capture the structural similarity of time series. However, one drawback of the bag-of-patterns representation is that its dimension may be very high, which limits its application for large datasets. For instance, when the size of the alphabet τ and the number of symbols w are 4 and 8, respectively, the dimension of the bag-of-patterns representation could reach $\tau^w = 65,536$.

In this work, motivated by the success of the bag-of-words model in text document analysis [16,17] and image analysis [18,19], we propose a simple yet effective bag-of-words representation whose dimension is much lower than the bag-of-patterns representation to characterize biomedical time series such as ECG and EEG signals. The bag-of-words representation is able to capture high-level structural information of time series due to the utilization of both local and global information. Furthermore, it can be used to process time series with different lengths attributing to the fact that it is constructed incrementally.

The bag-of-words model was originally developed for document representation. The basic idea is to define a codebook that contains a set of codewords and then represent a document as a histogram of the codewords, each entry of which is the count of a codeword occurred in the document. Although the order information of words is ignored, the bag-of-words model is still very effective to capture document information because the frequency information of codewords in documents are well explored. Recently, the bag-of-words model is extended to analyze images and videos in computer vision [18,19]. Local patches extracted from images or videos are treated as words and the codebook is constructed by clustering all the local patches in the training data. Similar to the extension of the bag-of-words representation in computer vision, we here extend the bag-of-words representation to characterize biomedical time series by regarding local segments extracted from time series as words and treat the time series as documents.

1.1. Overview of the proposed approach

In the bag-of-words representation, a time series is treated as a text document and local segments extracted from the time series as words. The general flowchart of the proposed method is demonstrated in Fig. 1. Firstly, we continuously slide a window with a pre-defined length along the time series to extract a group of local segments. Then, we extract a feature vector from each of the local segments using DWT. Next, similar to the bag-of-visual-words model in images and videos analysis [18,19], all local segments from the training time series are clustered by k-means clustering to create a codebook, i.e., the cluster centers are treated as codewords. Then, a local segment is assigned the codeword that has the minimum distance to the local segment, and the time series is represented as a histogram of the codewords. Finally, the bag-of-words representation is used as input for classification.

1.2. Contribution and organization

The main contribution of the paper is twofold: (i) a simple yet effective bag-of-words representation that is originally developed for text document representation is extended for analysis of biomedical time series such as ECG and EEG signals; (ii) a series of experiments was conducted to investigate the effectiveness and robustness of the bag-of-words representation for ECG and EEG signals classification.

The structure of the paper is organized as follows. In Section 2, the proposed method including the bag-of-words representation, distance measures and classification method is described. Section 3 describes the experimental setup. Experimental results are reported and analyzed in Section 4. Discussion and conclusion are given in Sections 5 and 6, respectively.

2. Proposed method

In this section, we describe the bag-of-words representation for biomedical time series classification. The bag-of-words representation ignores the temporal order of local segments within a time series and represents the time series as a histogram of codewords i.e., local segments. Several distance measures are then introduced for the histograms comparison.

2.1. Bag-of-words representation

We continuously slide a window with pre-defined length along a time series and extract a group of local segments from the time series. A feature vector is then extracted from each of the local segments using the DWT to characterize the local segment. All the local segments from the training data are clustered by the k-means to construct a codebook that contains a set of codewords, i.e., the cluster centers. Then, a local segment is assigned the codeword that has minimum distance with the local segment. The bag-of-words representation ignores the temporal order of local segments in a time series and represents the time series as a histogram of codewords.

2.1.1. Local segments extraction

A group of local segments are extracted from each time series by continuously sliding a window with pre-defined length along the time series. As local segments from different time series may be at different scales, all the local segments are normalized to zero mean and standard deviation. We transform a local segments into wavelet domain and use approximations wavelet coefficients of DWT as a feature vector to represent the local segment.

The wavelet transform that analyzes a signal at different frequency bands provides both accurate frequency information at low frequencies and time information at high frequencies, which are very important for biomedical signal analysis. The choice of wavelet function and the number of decomposition levels is of importance for the multiresolution decomposition. In this work, a single level DWT with order 3 Daubechies wavelet function (db3) is employed to decompose a local segment into approximations coefficients and detailed coefficients. Similar to the work in [14], we used the approximation coefficients as a feature vector to represent the local segment. We do not directly use the raw value of local segments as feature vectors due to the fact that features using the approximations coefficients not only are more robust to noise than features using raw segments but also have nearly half dimension of the local segments. Alternatively, we can also use any other features that are able to well represent the local segments, such as the features based on derivatives of signals [20]. Since our main motivation and focus in this paper is to extend the bag-of-words framework for biomedical time series analysis, we do not compare the performance of the detailed coefficients and other features to represent the local segments.

2.1.2. Codebook formulation

In the text document analysis, a codebook (vocabulary) is a set of pre-defined words, which are also called codewords. The bag-of-words method counts the number of each codeword that exists in a document and provides a document-level representation using a histogram of codewords. In image and video analysis, such codebook is generally created by performing clustering on a group of local patches from training data, i.e., the codewords are defined as the clustering centers. The codeword that is nearest to a local patch is then assigned to the local patch. The spatial and temporal

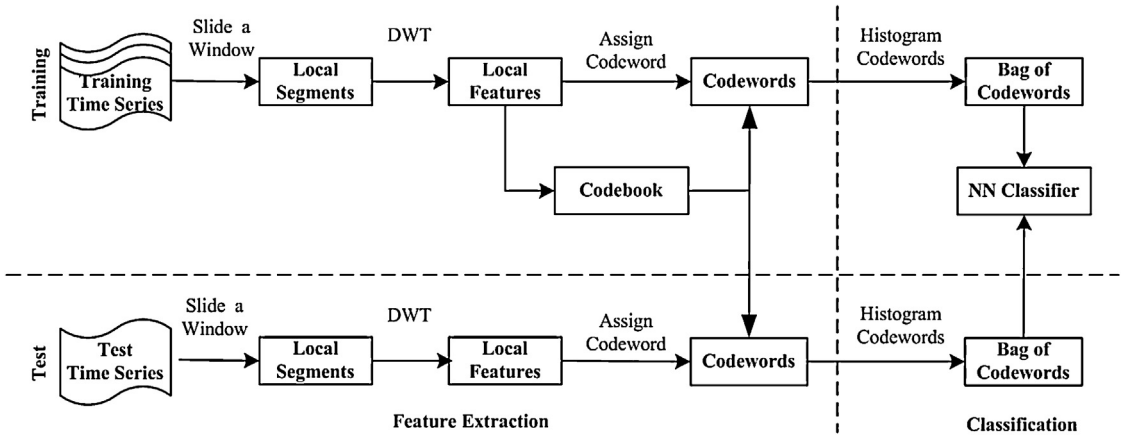

Fig. 1. The flowchart of the proposed bag-of-words approach for analysis of biomedical time series.

order information of local patches (codewords) is ignored and an image or video is represented as a histogram of codewords in the image or video. The classical k-means clustering algorithm [18,19] is commonly used to construct the codebook, although some other unsupervised and supervised methods are also developed such as mean-sift [21] and supervised Gaussian mixture models [22].

Similar to the codebook construction in image and video analysis, we cluster all the local segments from training time series using k-means clustering to construct the codebook. The clustering centers estimated by the k-means clustering are regarded as basis elements of the codebook, i.e., codewords. Suppose a group of local segments $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, where $\mathbf{x}_i \in \mathbb{R}^d$, are extracted from training time series, the codebook construction by k-means clustering is formulized as the optimization problem:

$$\min_{\mathbf{B}\in\mathbb{R}^{d\times K},\mathbf{V}\in\mathbb{R}^{K\times n}} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{B}\mathbf{v}_i\|_2, \quad \text{s.t.} \quad card(\mathbf{v}_i) = 1,\ |\mathbf{v}_i| = 1,\ \forall i,\ \mathbf{v}_i \geq 0, \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{d\times K}$ is the clustering centers and the vector $\mathbf{v}_i$ is the clustering index of the local segment $\mathbf{x}_i$, which is a unit-basis vector that has only one component equal to one and all the other components are zero. The codebook $\mathbf{B} \in \mathbb{R}^{d\times K}$ has K codewords, each of which is a d-length vector, the same length as the local segments. It is worth noting that the codebook only needs to be learned once from training data and it is universal for both training and test data.

The codebook size K is of importance to the bag-of-words representation. A compact codebook with too few entries has a limited discriminative ability, while a large codebook is likely to introduce noise due to the sparsity of the codewords histogram. Therefore, the size of the codebook should well balance the trade-off between discrimination and noise.

2.1.3. Codewords assignment

Once the codebook is constructed, a local segment is assigned the codeword that has minimum distance with the local segment. Specifically, suppose that a codebook with K entries, $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_K\}$, is learned from the training data. A local segment $\mathbf{x}_i$ is assigned the cth codeword that: $c^* = \underset{j}{argmin}\, d(\mathbf{b}_j, \mathbf{x}_i)$, where $d(\cdot, \cdot)$ is the Euclidean distance function.

After each local segment is assigned a codeword, the temporal order of local segments is ignored and a time series is represented as a histogram of codewords in the time series, each entry of which specifies the count of a codeword occurred in the time series. Fig. 2 illustrates the bag-of-words representation of an example EEG time series. The figure in the first row is the example EEG time series. The three figures in the second to fourth rows (left) are three local segments with length of 160 extracted from the time series, and the three figures in the second to fourth rows (right) are the corresponding codewords assigned to the local segments from codebook, which consists of 1000 codewords. The three local segments are assigned the 432nd, 118th, and 628th codewords, respectively. The figure in the last row is the bag-of-words representation for the time series, each entry of which gives the count of a codeword occurred in the time series.

2.2. Classifier

Some discriminative classifiers such as Artificial Neural Networks (ANN) [13], Support Vector Machine (SVM) [23], and Probabilistic Neural Networks (PNN) [14] are widely used for biomedical signal classification. In this paper, we use the simplest classifier, i.e., the 1-Nearest Neighbor (1-NN) classifier. Let $\mathbf{t}$ be a test time series and $\mathbf{R}^i$ represents the time series from the ith category. The test data is determined as the class C of the training sample that has minimal distance with the test data, i.e., $C^* = \underset{i}{argmin} D(\mathbf{t}, \mathbf{R}^i)$, where $D(\cdot, \cdot)$ is the similarity measure that is defined in the following.

It should be noted that the proposed bag-of-words representation is not limited to be used with the 1-NN classifier. The bag-of-words representation can be also input to some more promising classifiers such as the SVM, ANN and Decision Tree classifiers to improve the classification performance. Since our goal in this paper is to investigate the effectiveness of the bag-of-words representation, we used the simplest 1-NN for classification in the paper.

2.3. Similarity measure

Many similarity measures have been proposed for histograms comparison. In the following, we describe four commonly used similarity measures for distance measurement of two bag-of-words representations.

2.3.1. Euclidean distance

The Euclidean distance between histogram $\mathbf{h}$ and histogram $\mathbf{k}$ is defined as:

$$D_{L_2}(\mathbf{h}, \mathbf{k}) = \left(\sum_i \left| h(i) - k(i) \right|^2 \right)^{1/2}, \tag{2}$$

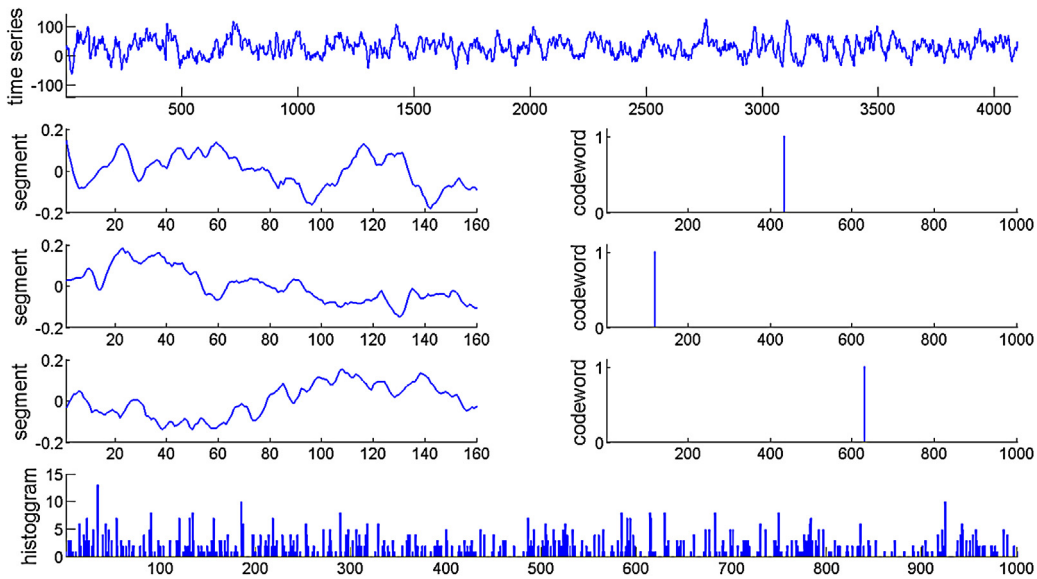

Fig. 2. The bag-of-words representation of an example time series. See the corresponding text for more details.

where $D_{L_2}(\mathbf{h}, \mathbf{k})$ is the Euclidean distance, which is commonly used in pattern recognition.

2.3.2. Chi-squared distance

The Euclidean distance subtracts the two histograms bin-by-bin and contributes each bin pairs equally to the distance. The problem is that some words such as "the", "but" and "however" occur more frequently in documents; therefore, they contribute more to the distance in the Euclidean distance measure. But they may actually have less discriminative information than rarely happened codewords. This leads to the Chi-squared distance (χ^2 distance):

$$D_{\chi^2}(\mathbf{h}, \mathbf{k}) = \sum_i \frac{\left|h(i) - k(i)\right|^2}{h(i) + k(i) + \varepsilon}, \tag{3}$$

where ε is a small value to avoid dividing by zero. The χ^2 distance introduces a normalization to emphasis the rarely happened codewords because common words are always shared between documents from different categories.

2.3.3. Jensen–Shannon distance

Each entry of the bag-of-words represents can be interpreted as the frequency of a codeword occurred in a time series. Therefore, the histogram stands for a probabilistic distribution over discrete random variables. A simple measure to compare two distribution is the Kullback–Leibler divergence:

$$D_{KL}(\mathbf{h}||\mathbf{k}) = \sum_i h(i)\left(\log_2 h(i) - \log_2 k(i)\right). \tag{4}$$

If and only if $\mathbf{h}$ and $\mathbf{k}$ are the same, the KL divergence becomes zero. In order to keep the distance symmetric, the Jensen–Shannon distance [24] is introduced as a symmetric extension:

$$D_{JS} = \frac{1}{2}\left(D_{KL}(\mathbf{h}||\mathbf{k}) + D_{KL}(\mathbf{k}||\mathbf{h})\right). \tag{5}$$

2.3.4. Histogram intersection based distance

The histogram intersection which counts the total overlap between two histograms is able to address the problem of partial matches when the two histograms have different sum over all the bins. The distance based on the histogram intersection is defined as [25]:

$$D_{HI}(\mathbf{h}||\mathbf{k}) = 1 - \sum_i \max\left(h(i), k(i)\right), \tag{6}$$

where $\mathbf{h}$ and $\mathbf{k}$ are normalized histogram vectors. Two histograms that have larger overlap will obtain a smaller distance.

2.4. Practical implementation

A large number of local segments may be extracted from training data, especially for large datasets. Clustering a large number of local segments to construct the codebook is computationally expensive. In practice, instead of using all the local segments extracted from the training data, we performing the k-means clustering on a subset of local segments randomly selected from the training data to construct the codebook. This strategy is also employed in image and video analysis to reduce the computation of codebook construction [18,19].

We continuously slide a window along a time series to extract local segments. However, when the time series contains too many data points, a large number of local segments will be extracted from the time series, which requires expensive computation. For instance, for a time series consisting of 2000 data points, about 1900 local segments will be extracted using a window with 100 length. The works in [26,27] extracted local segments by sliding random subwindows to reduce the number of extracted segments. Since we treat a time series as a document and local segments as words, too few segments will limit the discriminative ability of the bag-of-words. In practice, we can slide the window with a step of n data points ($n = 2$, 4, 6 or 8) along the time series to reduce the number of local segments extracted from the time series.

The MATLAB code of the bag-of-words representation in this work was made publicly available at http://www.mathworks.com/matlabcentral/fileexchange/38050.

3. Experimental setup

In this study, we applied the proposed bag-of-words representation to two biomedical applications, i.e., epilepsy detection from EEG signals and human identification from ECG signals. Three datasets constructed from EEG and ECG signals for epilepsy

Table 1
The three datasets used in the experiments.

Datasets	Classes	Num of signals	Length of signals
EEG	5	500	4096
ECG-40	40	2000	2048
ECG-15	15	1500	2048–4096

detection and human identification are used to evaluate the performance of the bag-of-words representation. The first dataset is collected from EEG signals and it is widely used for automatic epileptic seizure detection. The other two datasets are extracted from long ECG signals (more than 1,000,000 points) that collected from different subjects with random start points. Each of the long ECG signals corresponds to a class, i.e., subjects' identity.

3.1. EEG dataset for epilepsy detection

The EEG dataset described in [28] was widely used for automatic epileptic seizure detection. The complete EEG dataset consists of five classes (i.e., A, B, C, D, and E), each of which contains 100 single-channel EEG sequences of the same length 4096. All the signals were recorded with the same 128-channel amplifier system and visual inspected for artifacts. Set A and set B are collected from surface EEG recordings of five healthy subjects with eye open and eye closed, respectively. The other three sets (C, D and E) are taken from intracranial EEG recording of five patients suffered from epileptic. Set C and set D are taken from the epileptogenic zone and the hippocampal formation of the opposite hemisphere of the brain, respectively. Set C and set D were recorded in seizure-free intervals, whereas set E only contains seizure activity. Fig. 3 shows example time series from each of the five classes.

3.2. ECG-40 dataset for human identification

The ECG-40 dataset was obtained from the Fantasia ECG database [29], which consists of 40 healthy subjects. 40 long ECG signals are collected from each of the 40 subjects monitored for about 2 h with a sampling rate of 250 Hz. We extracted 50 time series of length 2048 from each of the 40 long signals consisting of more than 1,000,000 data points with random start points. Totally, the ECG-40 dataset contains 2000 time series of length 2048, which are evenly distributed in the 40 classes. The aim of the classification experiment is to assign a test ECG time series to the corresponding subject, i.e., the human identification.

3.3. ECG-15 dataset for human identification

The ECG-15 dataset consists of 1500 time series extracted from 15 long ECG signals in the BIDMC Congestive Heart Failure Database [29]. The 15 long ECG signals were recorded from 15 patients suffered from severe congestive heart failure. 100 time series of length between 2048 and 4096 are extracted from each of the 15 long ECG signals with random start points. Totally, the ECG-15 dataset consists of 15 classes, each of which has 100 time series of length between 2048 and 4096.

It is worth noting that although the extracted ECG time series in the same class are obtained from the same long ECG signal, there exist substantial inter-class variations. The aim of the ECG signal classification in our experiment is to attribute each instance, i.e., extracted ECG time series, to their subjects' identity, which can be used for human identification from ECG signals in real application.

Table 1 summaries the three datasets used in the experiments. It should be noted that the lengths of the 1500 time series in the ECG-15 dataset are not the same, which vary between 2048 and 4096.

4. Results

In this section, we report experimental results on the three datasets. Firstly, we investigated the impact of parameters by varying the length of local segments and the size of codebook K based on different distance measures. Then, we compared the proposed method with the Discrete Wavelet Transform (DWT) [14] representation, the Discrete Fourier Transform (DFT) [30] representation, the NN classifier based on Dynamic Time Warping (DTW) [31] distance and the bag-of-patterns representation (BoP) [15]. Next, we compared the epilepsy detection accuracies on the EEG dataset achieved by the proposed method with those achieved by other state-of-the-art methods. In addition, we compared the performance of the proposed bag-of-words representation for human identification from ECG signals with previous state-of-the-art methods. Finally, we investigated the robustness of the bag-of-words representation to noise. In order to ensure an un-biased evaluation, a dataset is randomly partitioned into ten subsets. Nine subsets are used for training while the remaining one is retained for test. The classification process is then repeated ten times with each of the ten subsets used exactly once as test data. The average accuracies and the standard deviations are reported for evaluation.

4.1. Length of local segments

We varied the length of local segments between 8 and 256 in the experiments. The determination of such parameter ranges relies on the fact that the biomedical time series such as ECG and EEG signals are relatively flat. The classification accuracies on the EEG, ECG-40 and ECG-15 datasets with a codebook size of 1000 using the Chi-squared distance is illustrated in Fig. 4(a)–(c), respectively. From the experimental results, it can be seen that the performance is relatively stable with respect to the length of local segments when it is between 64 and 192. The classification accuracies decrease considerably with the length less than 16. This is mainly due to the fact that a local segment with too short or too long length cannot capture local structure information within time series. In the following experiments, we empirically set the length of local segments as 128.

4.2. Codebook size

To show the performance of the bag-of-words representation with respect to the size of the codebook, we report the classification accuracies on the three datasets in Fig. 5, increasing the size of the codebook from 10 to 3500. We can see that the results become very stable when the size of the codebook is larger than 500. The classification accuracies reduce quickly if the size of the codebook is less than 100, which confirms that a compact codebook with too few entries has a limited discriminative ability. The optimal size of the codebook can be roughly identified as 1000–3500.

4.3. Distance measurement

We compared the classification performance on the three datasets using the four similarity measures described in Section 2.3. Fig. 6 demonstrates the classification accuracies based on the four distance measures with the codebook size of 10, 100, 1000 and 2000. We can see that the results are slightly different using various distance measures, indicating that the distance measures have limited impact on the performance of the bag-of-words representation. Overall, the Chi-squared distance measure performs slightly better than the other measures for all the four sizes of the codebook.

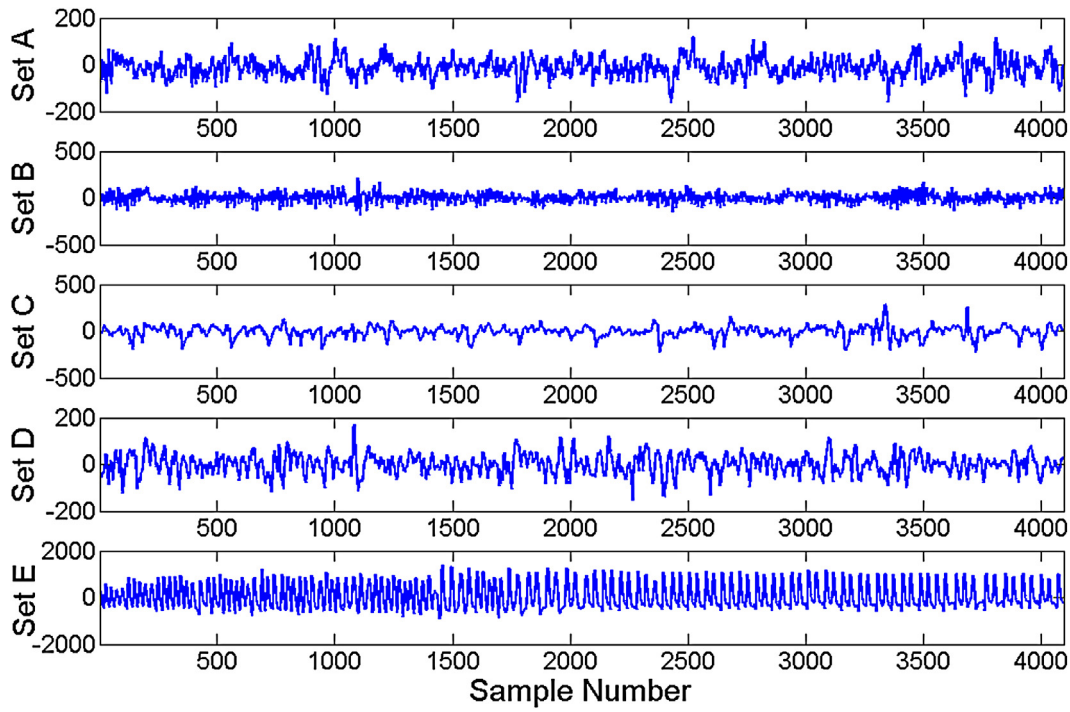

Fig. 3. Example EEG sequences from each of the five classes.

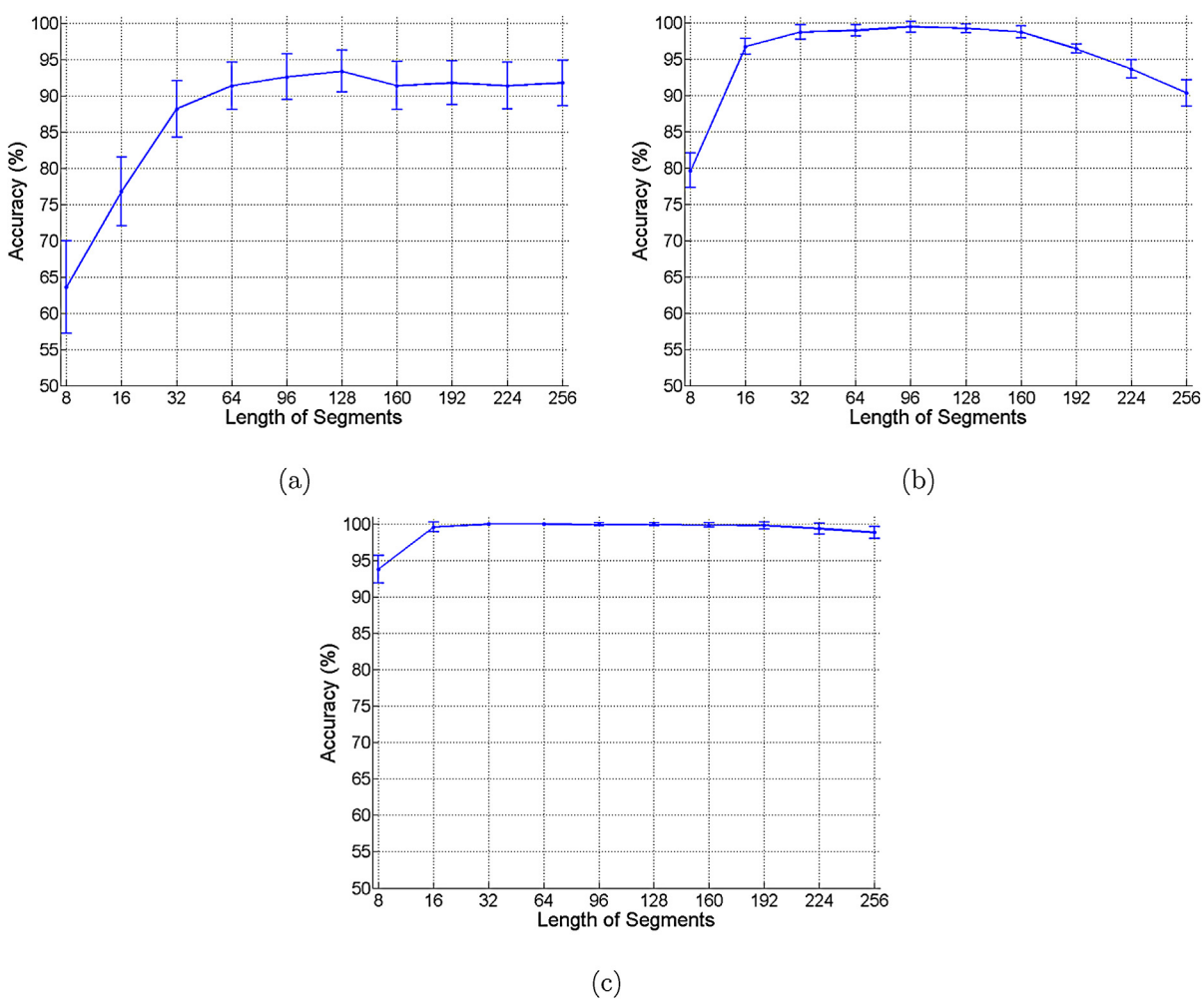

Fig. 4. Classification accuracies with respect to the length of segments on the EEG (a), ECG-40 (b) and ECG-15 (c) datasets, respectively.

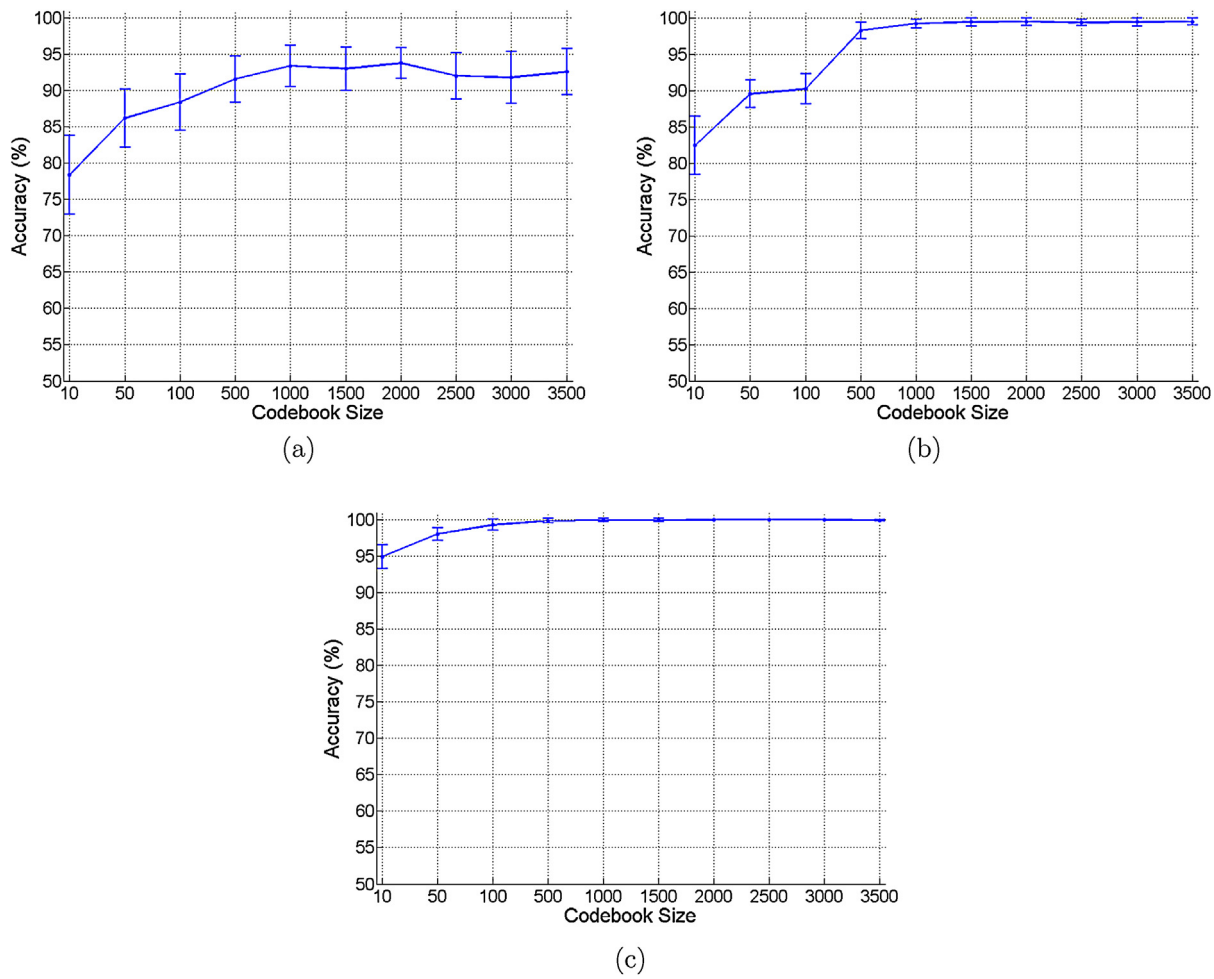

Fig. 5. Classification accuracies with respect to the codebook size on the EEG (a), ECG-40 (b) and ECG-15 (c) datasets, respectively.

4.4. Comparison with classical methods

We compared the performance of the proposed bag-of-words representation with that of the DWT representation [14], the DFT representation [30], and the NN classifier based on the DTW distance [31]. In addition, we also compared the proposed bag-of-words representation with the bag-of-patterns representation [15], which is very similar to the proposed approach.

- DWT that represents a signal in multiresolution is able to capture both frequency and location information of time series. Similar to the DWT based feature used in [14], we used the Daubechies wavelet (db2) and decomposed the time series into 4 levels. The detail wavelet coefficients of the four levels and the approximation wavelet coefficients of the fourth level are concatenated to form the final representation.
- DFT is a widely used transformation technique to extract frequency information from time series. We transformed the original time series into the frequency domain and extracted the DFT coefficients as features.
- DTW that uses dynamic programming technique to determine the best alignment of two sequences is able to deal with temporal drift between time series. The distance matrix of each pair of the test time series and the training time series is calculated based on the unconstrained DTW. This distance matrix is used as input of the NN classifier.
- The BoP representation that represents a time series as a histogram of local patterns is very similar to the proposed bag-of-words representation. The size of alphabet τ and the number of symbols w are empirically set to 4 and 6, respectively. We varied the length of local segments in the bag-of-patterns representation from 16 to 320 with a step of 16. The best accuracy is reported for comparison.

Since the time series in the ECG-15 datasets have different lengths (2048–4096), we resized all the time series to the same length of 4096 using bilinear interpolation so that the DWT and DFT based features have the same dimension. When calculating the DTW distance, we reduced all the time series in the three datasets to the length of 820 with a downsampling rate of about 5 because DTW is computationally expensive.

Table 2 summarizes the best results achieved by the proposed approach and the other methods. It can be seen that the proposed approach achieves the highest accuracies (93.8% on the EEG dataset, 99.5% on the ECG-40 dataset, and 100% on the ECG-15 dataset, respectively), which illustrate the effectiveness of the

Table 2
Comparison of results on the three datasets using different methods.

Methods	EEG	ECG-40	ECG-15
DWT	76.0 ± 4.8	25.1 ± 4.8	20.1 ± 4.4
DFT	91.6 ± 3.5	85.6 ± 2.3	60.6 ± 3.2
DTW	71.6 ± 5.3	74.5 ± 3.5	85.5 ± 2.7
BoP [15]	87.8 ± 2.3	99.4 ± 0.9	99.8 ± 0.3
Proposed method	93.8 ± 2.1	99.5 ± 0.5	100 ± 0

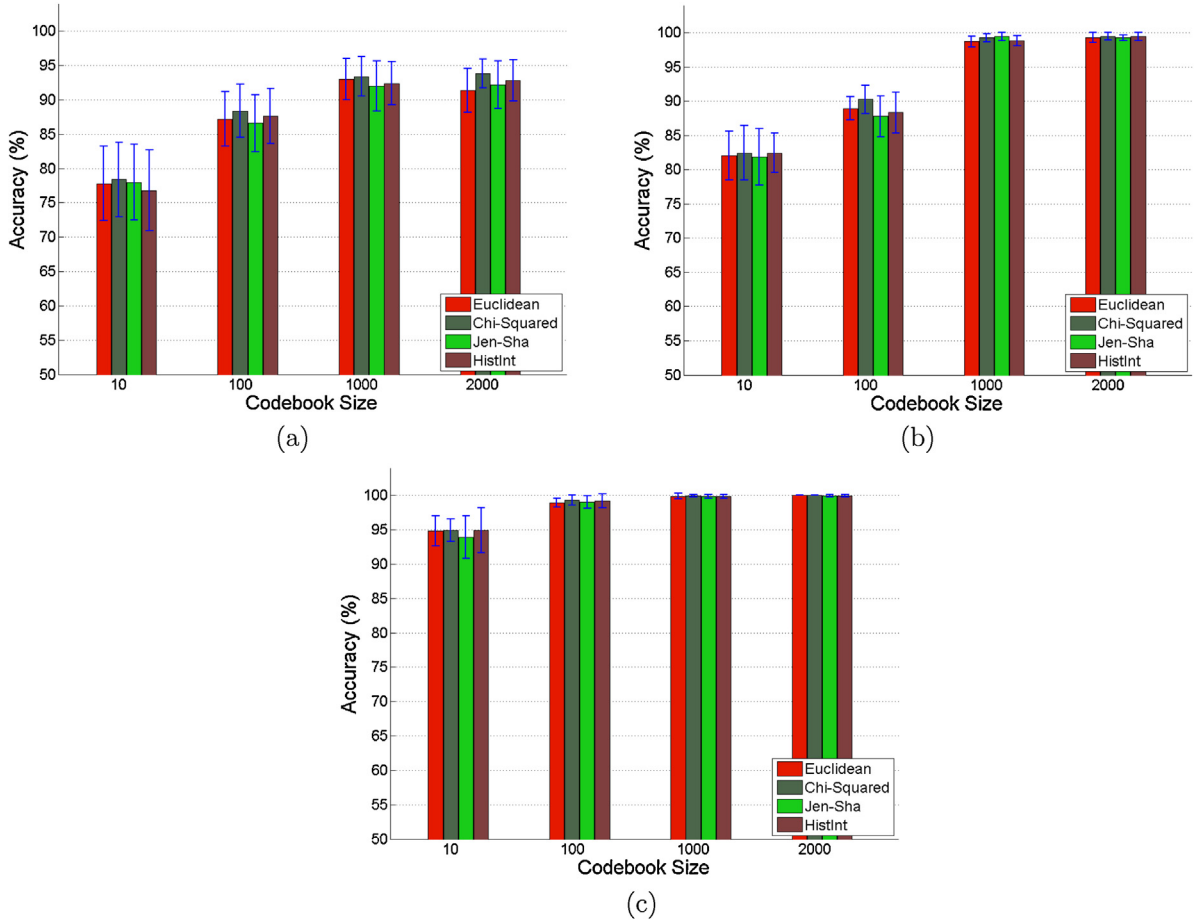

Fig. 6. Classification accuracies using different distance measures on the EEG (a), ECG-40 (b) and ECG-15 (c) datasets, respectively. The figure is best viewed in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

bag-of-words representation. The BoP representation obtains comparable accuracies on the ECG-40 and the ECG-15 datasets with that by the bag-of-words representation. However, the proposed bag-of-words representation performs significantly better than the BoP representation on the EEG dataset. The DFT feature and DTW distance methods outperform the DWT based method. This is probably because that the DFT and DTW can better deal with temporal sift between sequences than the DWT.

4.5. Comparison with previous epilepsy detection methods

The EEG dataset used in our experiment is a popular dataset for automatic epileptic seizure classification and localization. Table 3 provides a comparison of the classification accuracies between the proposed bag-of-words method and previous state-of-the-art approaches in the literature. It should be noticed that the comparison is not direct, since the aim of our method is to classify the time series at sequence level, while the other methods are to classify segments extracted from the time series. Some works used only several subsets of the whole EEG dataset to construct a 2-class dataset, while others used the whole EEG dataset with 5 classes.

From the table, we can see that the bag-of-words method outperforms most of the other methods. For the 5-class classification where the whole EEG dataset is used, the classification accuracies of support vector machine (SVM), probabilistic neural network (PNN) and multilayer perception neural network (MLPNN) with raw data are 75.60%, 72.00% and 68.80% [23], respectively. When features extracted from DWT and lyapunov exponents are used, the corresponding accuracies increase to 99.28%, 98.05% and 93.63% [23], respectively. The result obtained by the proposed bag-of-words representation with the simplest NN classifier is slightly lower than those achieved by SVM and PNN with features based on DWT and lyapunov exponents. However, it is slightly higher than the result obtained by MLPNN (93.63%) with features based on DWT and lyapunov exponents.

4.6. Comparison with previous human identification methods

Human identification from ECG signals has been extensively investigated in the literature [35,36,40,41,43]. Table 4 compares the proposed method with previous state-of-the-art methods for human identification from ECG signals.[1] It can be seen that the proposed method achieved comparable or higher accuracy comparing to the state-of-the-art methods.

One advantage of the proposed bag-of-words representation is that it does not require segmentation of individual heartbeats or detection of any fiducial points. By contrast, the other method except the one in [44] need to segment individual heartbeats or detect fiducial points to extract discriminative features, which is an arduous procedure, especially for ECG signals contaminated by noise. For instance, the method in [40] segmented individual heartbeats and applied the principal component analysis (PCA) for feature extraction. The method in [41] detected the R peak to retrieve ECG waveform for characterizing a whole ECG signal.

[1] Since most datasets and source code in the previous works are publicly unavailable, the comparison is not directly performed on the same dataset.

Table 3
Comparison of the classification accuracy (%) on the epileptic EEG dataset. In the "datasets" column, the data in the parentheses are in the same class, while the semicolon separates different classes.

Methods	Datasets	Class Num	Accuracy (%)
Spectral estimation + PNN [1]	(A, C); E	2	97.50
Entropies + fuzzy classifier [2]	(A, B, C, D); E	2	98.1
Entropy + neuro-fuzzy inference [3]	A; E	2	95
FFT + decision tree [32]	A; E	2	98.72
Nonlinear features + extreme learning [33]	D; E	2	96.00
DWT + approximate entropy [34]	(A, C, D); E	2	96.65
Line length features + ANN [13]	A; E	2	99.6
	(A, C, D); E	2	97.75
	(A, B, C, D); E	2	97.77
Raw data + SVM [23]	A; B; C; D; E	5	75.6
Raw data + PNN [23]	A; B; C; D; E	5	72.0
Raw data + MLPNN [23]	A; B; C; D; E	5	68.8
Lyapunov exponents + SVM [23]	A; B; C; D; E	5	99.3
Lyapunov exponents + PNN [23]	A; B; C; D; E	5	98.1
Lyapunov exponents + MLPNN [23]	A; B; C; D; E	5	93.6
Bag-of-words + 1-NN	A; E	2	99.5 ± 0.7
	(C, D); E	2	98.9 ± 0.9
	(A, C, D); E	2	99.0 ± 0.9
	(A, B, C, D); E	2	99.2 ± 0.8
	A; B; C; D; E	5	93.8 ± 2.1

Unlike these methods that are based on individual heartbeats or fiducial points, the proposed method directly operates on the ECG signals that contain several heartbeat waveforms and do not need to detect the fiducial points.

It is worth noting that the aim of the human identification from ECG signals is different from heartbeats classification [10,12] which has been extensively investigated. The aim of human identification is to assign a test ECG signals that may contain several individual heartbeats to a certain class (subject), while the heartbeats classification is to classify individual heartbeats, which needs to segment individual heartbeats from ECG signals. The previous methods for heartbeats classification is not applicable for the human identification task in the experiment.

Table 4
Human identification from ECG signals for comparison. SN stands for subject number. HS and FPD denote heartbeat segmentation and fiducial points detection, respectively.

Methods	SN	HS or FPD	Accuracy (%)
Biel et al. [35]	20	Yes	100
Israel et al. [36]	29	Yes	100
Wübbeler et al. [37]	74	Yes	98.1
Wang et al. [38]	13	Yes	100
Chan et al. [39]	50	Yes	95
Irvine et al. [40]	43	Yes	~100
Fang and Chan [41]	100	Yes	99
Zhao and Yang [42]	20	Yes	95.3
Pal and Mitra [43]	20	Yes	95
Plataniotis et al. [44]	14	No	100
Bag-of-words	**15**	**No**	**100 ± 0**
Bag-of-words	**40**	**No**	**99.5 ± 0.5**

Table 5
Classification accuracies (%) on the three datasets corrupted by zero-mean white Gaussian noise.

SNR	EEG	ECG-40	ECG-15
10 dB	92.6 ± 2.53	98.9 ± 0.82	99.8 ± 0.52
8 dB	91.8 ± 3.23	98.4 ± 1.02	99.7 ± 0.47
6 dB	91.2 ± 3.53	97.6 ± 1.09	99.6 ± 0.55
4 dB	90.4 ± 4.26	95.5 ± 1.36	99.2 ± 0.54
2 dB	88.8 ± 4.26	92.6 ± 1.89	98.9 ± 0.78
0 dB	85.2 ± 5.35	89.9 ± 2.54	98.6 ± 0.85

4.7. Robustness to noise

This experiment is designed to investigate the robustness of the bag-of-words representation to noise. All signals in the EEG, ECG-40 and ECG-15 datasets were corrupted by zero mean white Gaussian noise. The standard deviation of the white Gaussian noise is varied so that the SNRs are between 10 dB and 0 dB. The training data and the test data are separated exactly the same as those in the previous experiments. Table 5 summaries the classification accuracies on the three datasets contaminated by the white Gaussian noise with different SNRs. It can be seen that the bag-of-words approach is relatively robust to noise. The accuracies decreased by less than 2% when the SNR is 10 dB. Even for considerable noise contamination with the SNR 0 dB, the accuracies reduced less than 10% for the EEG and ECG-40 datasets, and only less than 2% for the ECG-15 dataset.

5. Discussion

Fig. 7 shows four time series (left) and the corresponding bag-of-words representation (right). The first two time series are from the same class (class A) while the third and the fourth time series are from different classes (class C and class D, respectively). As can be seen, the codewords distribution of the first two time series that belong to the same class are very similar while the codewords corresponding to the other two time series that belong to different classes are more different. This is probably due to the fact that local segments extracted from time series that belong to the same classes have similar structures and they are projected to similar codewords. Therefore, these time series have similar codewords frequency significantly different to those corresponding to the other classes. The codewords distribution of the first two time series are more similar that the others, which demonstrates that the bag-of-words representation is very discriminative.

The proposed bag-of-words representation is designed as a universal feature extraction method for biomedical time series. It is not limited to extract feature from EEG and ECG signals. The bag-of-words representation can be also extended to characterize some other kinds of biomedical time series such as Electromyography (EMG) signals and accelerometer signals. However, since the local segments are extracted by sliding a window along time series, the bag-of-words representation is ineffective for time series that is not reasonably long. This is mainly due to the limitation that the bag-of-words representation cannot extract enough meaningful and discriminative local segments from short sequences.

Another limitation of the proposed bag-of-words representation is that it fails to analyze the temporal order of local points/segments in a time series because it ignores the temporal order of local segments in a time series. Therefore, the bag-of-words representation is not applicable to the tasks that emphasize on local temporal information of local points/segments such as motif detection in time series [45,46] and ECG heartbeats classification [10,12].

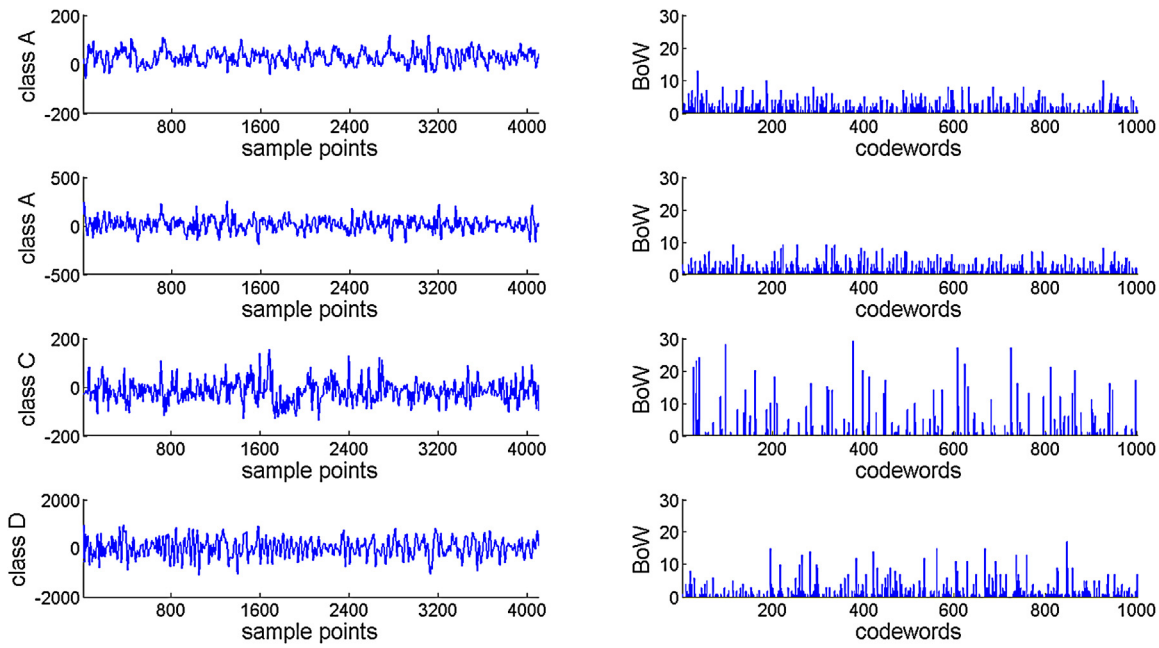

Fig. 7. Four time series (left) from the same class and different classes, and the corresponding bag-of-words representation (right). BoW stands for bag-of-words representation. It should be noted that this figure is an example showing the bag-of-words features belonging to the same and different classes, but not a comprehensive comparison of the features. See the corresponding text for more details.

By contrast, the advantage of the proposed bag-of-words representation is that it is able to effectively capture high-level structural information of a whole time series. The bag-of-words representation is an effective and efficient method to extract discriminative features to characterize a whole signal, but not individual segments in the signal.

The size of the codebook N is pre-defined and empirically determined in the method. A compact codebook with small size has a limited discriminative ability, while a codebook with large size is likely to introduce noise. How to adaptively set the optimal size of the codebook to make the codebook compact and yet discriminative is still an open question. Some criteria can be defined to merge entries of a codebook to construct an adaptive codebook. For instance, the method in [47] utilized Maximization of Mutual Information (MMI) principal to estimate the optimal N. Two entries of a codebook are merged by maximizing the mutual information in an unsupervised way. Creating a codebook with adaptive size will be investigated in our future work.

6. Conclusion

In this paper, we proposed a bag-of-words representation for biomedical time series analysis. The proposed method treats a time series as a document and local segments extracted from the time series as words. The time series is represented as a histogram of codewords. Although the temporal order information of the local segments is ignored, both local structure and global structure information of the time series are captured. Experimental results on three publicly available datasets demonstrate that the bag-of-words representation is effective for characterizing biomedical time series such as EEG and ECG signals. Furthermore, the bag-of-words representation is not only insensitive to the model parameters such as the length of local segments and the size of codebook, but also robust to noise.

We compared the performance of the bag-of-words representation with several state-of-the-art approaches for the task of epilepsy detection in EEG signals and human identification from ECG signals. The bag-of-words representation with the simplest 1-Nearest Neighbor (1-NN) classifier achieves comparable or higher classification accuracies than those by the others, which demonstrates that the bag-of-words representation is effective to characterize biomedical time series such as the EEG signals and ECG signals.

References

[1] A.R. Naghsh-Nilchi, M. Aghashahi, Epilepsy seizure detection using eigensystem spectral estimation and multiple layer perceptron neural network, Biomedical Signal Processing and Control 5 (2) (2010) 147–157.
[2] U.R. Acharya, F. Molinari, S.V. Sree, S. Chattopadhyay, K.-H. Ng, J.S. Suri, Automated diagnosis of epileptic EEG using entropies, Biomedical Signal Processing and Control 7 (4) (2012) 401–408.
[3] N. Kannathal, M.L. Choo, U.R. Acharya, P. Sadasivan, Entropies for detection of epilepsy in EEG, Computer Methods and Programs in Biomedicine 80 (3) (2005) 187–194.
[4] H. Wang, J. Xu, Local discriminative spatial patterns for movement-related potentials-based EEG classification, Biomedical Signal Processing and Control 6 (4) (2011) 427–431.
[5] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain computer interfaces for communication and control, Clinical Neurophysiology 113 (6) (2002) 767–791.
[6] H. Wang, Multiclass filters by a weighted pairwise criterion for EEG single-trial classification, IEEE Transactions on Biomedical Engineering 58 (5) (2011) 1412–1420.
[7] K.-Q. Shen, C.-J. Ong, X.-P. Li, Z. Hui, E. Wilder-Smith, A feature selection method for multilevel mental fatigue EEG classification, IEEE Transactions on Biomedical Engineering 54 (7) (2007) 1231–1237.
[8] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, J.-H. Chen, EEG-based emotion recognition in music listening, IEEE Transactions on Biomedical Engineering 57 (7) (2010) 1798–1806.
[9] T. Ince, S. Kiranyaz, M. Gabbouj, A generic and robust system for automated patient-specific classification of ECG signals, IEEE Transactions on Biomedical Engineering 56 (5) (2009) 1415–1426.
[10] A. Kampouraki, G. Manis, C. Nikou, Heartbeat time series classification with support vector machines, IEEE Transactions on Information Technology in Biomedicine 13 (4) (2009) 512–518.
[11] M. Huken, P. Stagge, Recurrent neural networks for time series classification, Neurocomputing 50 (2003) 223–235.
[12] A.E. Zadeh, A. Khazaee, V. Ranaee, Classification of the electrocardiogram signals using supervised classifiers and efficient features, Computer Methods and Programs in Biomedicine 99 (2) (2010) 179–194.
[13] L. Guo, D. Rivero, J. Dorado, J.R. Rabu-al, A. Pazos, Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks, Journal of Neuroscience Methods 191 (1) (2010) 101–109.

[14] İ. Güler, E.D. Übeyli, ECG beat classifier designed by combined neural network model, Pattern Recognition 38 (2) (2005) 199–208.
[15] J. Lin, R. Khade, Y. Li, Rotation-invariant similarity in time series using bag-of-patterns representation, Journal of Intelligent Information Systems (2012) 1–29.
[16] G. Lebanon, Y. Mao, J. Dillon, The locally weighted bag of words framework for document representation, Journal of Machine Learning Research 8 (2007) 2405–2441.
[17] D. Blei, A. Ng, M. Jordan, Latent Dirichlet allocation, Journal of Machine Learning Research 3 (2003) 993–1022.
[18] L. Fei-Fei, P. Perona, A Bayesian hierarchical model for learning natural scene categories, in: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2005, pp. 524–531.
[19] J.C. Niebles, H. Wang, L. Fei-Fei, Unsupervised learning of human action categories using spatial-temporal words, International Journal of Computer Vision 79 (3) (2008) 299–318.
[20] F. Scalzo, X. Hu, Semi-supervised detection of intracranial pressure alarms using waveform dynamics, Physiological Measurement 34 (4) (2013) 465–478.
[21] F. Jurie, B. Triggs, Creating efficient codebooks for visual recognition, in: Proc. IEEE Int'l Conf. Computer Vision, vol. 1, 2005, pp. 604–610.
[22] B. Fernando, E. Fromont, D. Muselet, M. Sebban, Supervised learning of Gaussian mixture models for visual vocabulary generation, Pattern Recognition 45 (2) (2012) 897–907.
[23] İ. Güler, E.D. Übeyli, Multiclass support vector machines for EEG-signals classification, IEEE Transactions on Information Technology in Biomedicine 11 (2) (2007) 117–126.
[24] D. Endres, J. Schindelin, A new metric for probability distributions, IEEE Transactions on Information Theory 49 (7) (2003) 1858–1860.
[25] K. Grauman, T. Darrell, The pyramid match kernel: efficient learning with sets of features, Journal of Machine Learning Research 8 (2007) 725–760.
[26] F. Scalzo, R. Hamilton, S. Asgari, S. Kim, X. Hu, Intracranial hypertension prediction using extremely randomized decision trees, Medical Engineering & Physics 34 (8) (2012) 1058–1065.
[27] F. Scalzo, P. Xu, M. Bergsneider, X. Hu, Random subwindows for robust peak recognition in intracranial pressure signals, in: Advances in Visual Computing, Lecture Notes in Computer Science, 2008, pp. 370–380.
[28] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, Physical Review E 64 (6 Pt 1) (2001).
[29] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) 215–220.
[30] D. Rafiei, A. Mendelzon, Querying time series data based on similarity, IEEE Transactions on Knowledge and Data Engineering 12 (5) (2000) 675–693.
[31] T. Chung Fu, A review on time series data mining, Engineering Applications of Artificial Intelligence 24 (1) (2011) 164–181.
[32] K. Polat, S. Günes, Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform, Applied Mathematics and Computation 187 (2) (2007) 1017–1026.
[33] Q. Yuan, W. Zhou, S. Li, D. Cai, Epileptic EEG classification based on extreme learning machine and nonlinear features, Epilepsy Research 96 (1/2) (2011) 29–38.
[34] H. Ocak, Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy, Expert Systems with Applications 36 (2 Pt 1) (2009) 2027–2036.
[35] L. Biel, O. Pettersson, L. Philipson, P. Wide, ECG analysis: a new approach in human identification, IEEE Transactions on Instrumentation and Measurement 50 (3) (2001) 808–812.
[36] S.A. Israel, J.M. Irvine, A. Cheng, M.D. Wiederhold, B.K. Wiederhold, ECG to identify individuals, Pattern Recognition 38 (1) (2005) 133–142.
[37] G. Wübbeler, M. Stavridis, D. Kreiseler, R.-D. Bousseljot, C. Elster, Verification of humans using the electrocardiogram, Pattern Recognition Letters 28 (10) (2007) 1172–1175.
[38] Y. Wang, F. Agrafioti, D. Hatzinakos, K.N. Plataniotis, Analysis of human electrocardiogram for biometric recognition, EURASIP Journal on Advances in Signal Processing 2008 (19) (2008).
[39] A. Chan, M. Hamdy, A. Badre, V. Badee, Wavelet distance measure for person identification using electrocardiograms, IEEE Transactions on Instrumentation and Measurement 57 (2) (2008) 248–253.
[40] J.M. Irvine, S.A. Israel, W.T. Scruggs, W.J. Worek, Eigenpulse: robust human identification from cardiovascular function, Pattern Recognition 41 (11) (2008) 3427–3435.
[41] S.-C. Fang, H.-L. Chan, Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space, Pattern Recognition 42 (9) (2009) 1824–1831.
[42] Z. Zhao, L. Yang, ECG identification based on matching pursuit, in: Proc. IEEE BMEI, vol. 2, 2011, pp. 721–724, http://dx.doi.org/10.1109/BMEI.2011.6098470.
[43] S. Pal, M. Mitra, Increasing the accuracy of ECG based biometric analysis by data modelling, Measurement 45 (7) (2012) 1927–1932.
[44] K. Plataniotis, D. Hatzinakos, J. Lee, ECG biometric recognition without fiducial detection, in: Proc. IEEE BCC, 2006, pp. 1–6.
[45] J. Lin, E. Keogh, P. Patel, S. Lonardi, Finding motifs in time series, in: Proc. of the 2nd Workshop on Temporal Data Mining, at the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002, pp. 53–68.
[46] P. Patel, E. Keogh, J. Lin, S. Lonardi, Mining motifs in massive time series databases, in: Proc. IEEE Int'l Conf. on Data Mining, 2002, pp. 370–377.
[47] J. Liu, M. Shah, Learning human actions via information maximization, in: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2008, pp. 1–8.