KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Moeyersons J., Varon C., Testelmans D., Buyse B., Van Huffel S. (2017), **ECG Artefact Detection Using Ensemble Decision Trees** In Proceedings of the 44th Annual Computing in Cardiology Conference - CinC, 2017, Rennes, France
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	Na.
Journal homepage	https://www.cinc2017.org/
Author contact	Jonathan.moeyersons@kuleuven.be 0479427028
Abstract	
IR	Na.

ECG Artefact Detection Using Ensemble Decision Trees

Jonathan Moeyersons[1,2], Carolina Varon[1,2], Dries Testelmans[3], Bertien Buyse[3], Sabine Van Huffel[1,2]

[1] KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium
[2] Imec, Leuven, Belgium
[3] UZ Leuven, Department of Pneumology, Leuven, Belgium

Abstract

This paper describes a novel method for artefact detection in electrocardiogram (ECG) signals.

ECG analysis algorithms require a relatively clean dataset. Therefore, data corrupted by artefacts should either be filtered or discarded. The proposed method can be situated in the second class, since it identifies contaminated segments that can later be discarded from further analysis.

The dataset used in this study contains 16 single lead ECG recordings, segmented in intervals of 60 seconds. Each segment is labeled either clean or contaminated by a medical doctor. Only 3.2% of the data is contaminated.

The segments are characterized by features derived from their autocorrelation function (ACF). Due to its effectiveness in skewed datasets, the RUSBoost algorithm is then used for classification. Results show an accuracy of 99.85%, a sensitivity of 100% and a specificity of 95.51%. This suggests that the proposed method could be of great help for future ECG processing.

1. Introduction

The electrocardiogram (ECG) is a fundamental tool in screening programs for cardiac health. The cost-effectiveness of this method has been shown in all age groups: adults, young athletes, neonates,… (1). However, due to the briefness of this test, it is possible that certain heart rhythm abnormalities are not detected. This lack of continuous measurement initiated a shift towards ambulatory monitoring devices. These devices allow monitoring for a longer period, thereby increasing the likelihood of detecting abnormalities.

Due to the ambulatory nature of these devices, motion artefacts, loose electrodes and interference from other electrical devices could cause distortions of the signal. The presence of these distortions, further referred to as artefacts, could reduce the diagnostic capabilities of the monitoring device and lead to inappropriate treatment decisions.

Different methodologies have been proposed to detect artefacts and enhance signal quality. One of these methods is blind source separation, e.g. ICA (2). This method acts on the whole signal and results in an enhancement of the overall quality of the ECG signal. However, no information on the artefact location can be extracted with this method. A frequently used way to overcome this problem is to segment the signal. This is typically followed by a feature extraction and classification step. Spectral and statistical information are often used as features (3,4). Furthermore, a variety of machine learning techniques has been used for classification of the segments. The downside of this methodology is the need for a gold standard.

A novel method is proposed to automatically identify the location of artefacts in long-term ECG recordings. The method starts by segmenting the ECG signal and characterizing each segment by its autocorrelation function (ACF). Features derived hereof are fed to a RUSBoost algorithm for classification. The choice for the ACF is motivated by the fact that it takes advantage of the repetitiveness of the ECG signal. Furthermore, it has been shown that the ACF of a clean segment is significantly different from a contaminated one (5).

This paper presents the results of the automatic classification of ECG segments using ACF features and a hybrid algorithm, RUSBoost. The performance of this classification is compared with the performance of a different algorithm obtained on the same dataset (5).

2. Materials and methods

This section contains a description of the dataset, the processing and a detailed explanation of the implementation of the RUSBoost algorithm. All analysis were performed in Matlab.

A. Data

The dataset used in this study contains 16 single lead ECG recordings of 16 different patients from the sleep laboratory of the University Hospital Leuven (UZ Leuven), Belgium. A total amount of 152h and 12min of

ECG was acquired at a sampling frequency of 200Hz.

Each recording was manually labeled by a medical doctor with adequate labeling experience. Segments were divided in two classes: Clean or contaminated. In total, 147h 18 min (96.8%) was labeled clean and 4h 54min (3.2%) contaminated. This skewness is considered for the choice of classification algorithm, which will be explained later. The proposed artefact detection method consists of two main steps detailed in the next subsections.

B. Segmentation, pre-processing and feature extraction

The presented dataset was originally intended for sleep apnea classification. In that research field, it is customary to segment each signal first and then perform the analysis on a minute-by-minute basis (6). Taken this into consideration, each ECG recording was first segmented in intervals of 60 seconds.

As explained in the introduction, motion artefacts, loose electrodes and interference from other electrical devices cause distortions of the signal.

The presence of electrical devices could cause power line interference. This is a narrow-band signal centered around 50 or 60Hz. Furthermore, due to breathing and patient movement baseline wander could occur. Both contaminating factors could influence the signal, but can easily be removed by the correct filter. Therefore, the segmented signal is filtered by means of a zero phase, band pass Butterworth filter with cut-off frequencies of 1 and 40Hz. This removes the baseline wander and power line interference. Additionally, the mean of each segment is subtracted to align the signal with the zero line.

In theory, it is possible to derive features from the one minute segments and build a classifier based hereon. However, research by Varon et al. has shown that a window reduction causes an improvement of the algorithm's resolution (5). Moreover, visual analysis of the noisy segments showed that often only small parts of the segments are contaminated. Therefore, it was opted to additionally segment each 60 seconds interval in intervals of 5 seconds.

Of every small segment, the ACF is computed with a maximum lag of 250ms. This depicts the repetitiveness of the ECG signal, but avoids inclusion of two consecutive R-peaks. As mentioned in the introduction, a clear change in the ACF can be observed if an ECG is contaminated by an artefact. In Figure 1, one can observe the difference between the ACF of a clean and a contaminated segment.

From the ACF, different features can be derived, among which the time lag of the different peaks and valleys, the amplitudes at different time lags and the similarity of the different ACF functions in a one minute interval.

The first feature that is selected from the 5 second intervals is the minimum time lag of the first saddle point of the ACF. In a clean segment, this time lag coincides with a shift of the R-peak towards the deepest point of the S-wave. So this value represents the duration of the RS interval. Excessive lengthening or shortening of this interval indicates the presence of artefacts. However, this saddle point is not always present due to e.g. a lack of S-wave. Therefore, as a second feature the maximum amplitude at the estimated time lag of the first saddle point of an average heartbeat, located at 35ms, is proposed. The third and final feature that is selected, is a measure of the similarity of the different ACFs. For every 60s interval, the maximum range between the different ACFs of the smaller segments, in a time lag interval between 30ms and 115ms, is computed. A large range indicates a lack of similarity between the different ACFs within that interval.

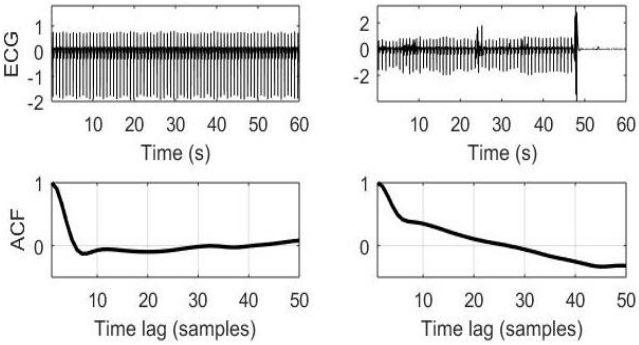

Figure 1: The ACF (bottom) of two different one minute segments (top). The first one corresponds to a clean segment, whilst the second corresponds to a contaminated segment. A clear difference between both ACF's can be observed.

The features were divided in a training and test set (70% / 30%). The training set is selected using the fixed-size algorithm, which maximizes the Renyi entropy (7). This ensures that, instead of having a random distribution, the training set distribution approximates the underlying distribution of the entire dataset. The training set is further used to train the artefact classification model.

C. Classification

In the case of a skewed dataset, traditional classification algorithms tend to classify the minority class to the majority class. This often results in a high overall accuracy but, an inaccurate classification of the minority class. In this dataset, the clean segments greatly outnumber the contaminated. Therefore, implementation of a classification algorithm that is able to effectively identify the rarely occurring contaminated segments is required.

Several techniques have been proposed to overcome the problem of class imbalance, including data sampling and boosting. Data sampling, on the one hand, balances the class distribution either by oversampling the minority class or under sampling the majority class. Boosting, on the other hand, can improve the performance of any weak classifier by iteratively building the classifier. Seifert et al, combined random under sampling (RUS) and boosting into

a new, hybrid, ensemble classification algorithm named RUSBoost (8). In the following, both components will be explained in more detail.

- RUS: The desired class balance is achieved by randomly removing samples from the majority class. The main drawback of this method is the loss of information due to the training sample deletion. On the upside, the time required to train the model is decreased.
- Boosting: This method builds a strong, ensemble classifier by making a linear combination of weak classifiers. The most common boosting algorithm is AdaBoost (9), which iteratively builds an ensemble classifier. Boosting can be applied on various weak learners, but in this study, it was opted to use decision trees.

By combining both methods, the main drawback of RUS, the loss of information, is overcome. The complete outline of the method can be found in (8).

Deep trees, with a minimal leaf size of 5 were used for higher ensemble accuracy. The training of each decision tree is established by the CART algorithm. The learn rate was set at 0.1 to additionally increase the accuracy.

The performance of this model is evaluated by computing the accuracy, sensitivity, specificity and balanced accuracy. The latter being the average of the sensitivity and specificity in order to take the skewness of the dataset into account (10).

3. Results and discussion

A. Classifier

The initial number of iterations was fixed at 1500. Since these settings create a large ensemble, the model needs to be compacted. To do so, the mean squared classification error of the training set is used as a determining factor.

In Figure 2, the classification error is plotted against the number of weak learners. A clear elbow in the classification errors can be observed at 100 weak learners. Therefore, weak learners 101 until 1500 are removed from the model. The final model consists of 100 weak learners.

B. Performance

The total number of segments under investigation is 9132 and only 294 of them are manually labeled as contaminated. This labelling allows to validate the proposed algorithm and compare the obtained performance with the performance of a previously developed algorithm on the same dataset (5).

As explained in section 2, first the intervals are segmented in smaller intervals of 5 seconds. From these 5 second intervals, the ACF features are derived. They are

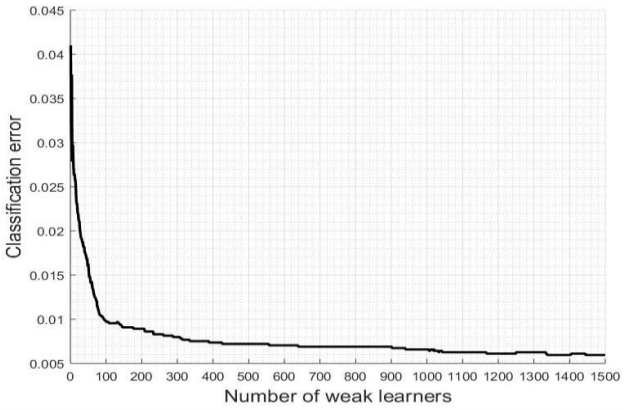

Figure 2: The classification error of the training set versus the number of weak learners. An elbow can be seen at approximately 100 weak learners.

fed to the RUSBoost algorithm which classifies them in two groups: Clean or contaminated. The performance of the proposed algorithm can be observed in Table 1.

In order to allow comparison between the novel and the previously developed method, further referred to as the 95th percent method, we applied it on the same test set. The results of this experiment are also displayed in Table 1.

It can be observed that the novel algorithm outperforms the other, because both accuracy, sensitivity, specificity and balanced accuracy are higher when computed on the test set.

Two reasons might explain these results. First, the 95th percent rule assumes that 5 percent of the data is contaminated, no matter the input. This assumption could never be completely fulfilled for this dataset, since only 3.2% of the data is labeled as contaminated. The RUSBoost algorithm does not make such assumption and might therefore perform better.

A second reason might be the effectiveness of the proposed algorithm on the presented, skewed dataset. A better accuracy, than the 95th percent algorithm, can already be observed, but a bigger difference can be observed in the balanced accuracy. This is a much more effective performance measure when dealing with skewed datasets. It can be observed that the balanced accuracy of the proposed method is 8.53% higher, compared to that of the 95th percent method. This is an additional indication that the RUSBoost algorithm is effective on skewed datasets.

Table 1: Comparison of the performance of the novel and the 95th percent method, developed in (5). Clear improvements can be observed in all performance metrics.

	RUSBoost	**95th**
Accuracy	99.85%	97.01%
Sensitivity	100%	97.55%
Specificity	95.51%	80.90%
Balanced Accuracy	97.75%	89.22%

4. Conclusion

A novel algorithm to detect artefacts in ECG signals by means of ACF features and the RUSBoost algorithm is presented in this paper.

Two advantages can be observed using this methodology. The first advantage is that it is capable of detecting artefacts in an online fashion. No comparison between different segments is made, therefore each segment can be analyzed separately. This allows a more efficient analysis of ECG segments.

The ability to locate contaminated segments in the ECG signal is the second advantage. This is in contrast with methods that enhance the overall quality of the signal, such as ICA.

A disadvantage of the method is the type of dataset on which it is trained. The dataset contains ECG signals from sleeping people. Therefore, it probably does not contain the same, or as much, contaminating factors as during daytime. Due to this, it might be possible that the algorithm performs worse on a daytime dataset.

Overall, the performance of the proposed algorithm is very good on this dataset. In future work, it is necessary to test the model on different datasets, containing more artefacts, to define the actual value for real life artefact detection. Furthermore, the class probability could also be investigated for a continuous quality indication of each ECG segment.

Acknowledgements

SV: Bijzonder Onderzoeksfonds KU Leuven (BOF): SPARKLE #: IDO-10-0358, The effect of perinatal stress on the later outcome in preterm babies #: C24/15/036, TARGID #: C32-16-00364; Agentschap Innoveren & Ondernemen (VLAIO): Project #: STW 150466 OSA + O&O HBC 2016 0184 eWatch; iMinds Medical Information Technologies: SBO-2016, ICON: HBC.2016.0167 SeizeIT; European Research Council: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (n° 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

References

1. Quaglini S, Rognoni C, Spazzolini C, Priori SG, Mannarino S, Schwartz PJ. Cost-effectiveness of neonatal ECG screening for the long QT syndrome. 2006;1824–32.
2. Chawla MPS, Verma H., Kumar V. Artifacts and noise removal in electrocardiograms using independent component analysis. Int J Cardiol. 2008;129(2):278–81.
3. Zaunseder S, Huhle R, Malberg H. CinC Challenge - Assessing the Usability of ECG by Ensemble Decision Trees. 2011;277–80.
4. Chudacek V, Zach L, Kuzilek J, Spilka J, Lhotska L. Simple scoring system for ECG quality assessment on Android platform. 2011 Comput Cardiol. 2011;449–51.
5. Varon C, Testelmans D, Buyse B, Suykens JAK, Van Huffel S. Robust artefact detection in long-term ECG recordings based on autocorrelation function similarity and percentile analysis. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2012. p. 3151–4.
6. Zywietz CW, Einem V von, Widiger B, Joseph G. ECG Analysis For Sleep Apnea Detection. Vol. 43, Methods Archive. F.K. Schattauer; 2004. 56-59 p.
7. De Brabanter K, De Brabanter J, Suykens JAK, De Moor B. Optimized fixed-size kernel models for large data sets. Comput Stat Data Anal. 2010;54(6):1484–504.
8. Seiffert C, Khoshgoftaar TM, Hulse J Van, Napolitano A. RUSBoost : A Hybrid Approach to Alleviating Class Imbalance. 2010;40(1):185–97.
9. Freund Y, Schapire R. A desicion-theoretic generalization of on-line learning and an application to boosting. Comput Learn theory. 1995;55:119–39.
10. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. Proc - Int Conf Pattern Recognit. 2010;3121–4.