Biomedical Signal Quality Assessment via Learning to Rank with an Application to Mechanical Heart Signals

Olli Lahdenoja[1], Mojtaba Jafari Tadi[1,2], Matti Kaisti[1], Timo Knuutila[1], Mikko Pänkäälä[1], Tero Koivisto[1]

[1] Department of Future Technologies, University of Turku, Finland
[2] Faculty of Medicine, University of Turku, Finland

Abstract

Traditionally the machine learning assisted quality assessment of biomedical signals (such as electrocardiogram - ECG, photoplethysmography - PPG) have classified a signal segment quality as "good" or "bad" and used this assessment to determine if the segment is usable for further processing steps, such as heart beat estimation. In principle, this is a suitable approach and can be justified by its straightforward implementation and applicability. However, in the case of body sensor networks with multiple simultaneously operating units, such as IMUs (Inertial Measurement Units) there is a need to select the best performing axes for further processing, instead of processing the data among all axes (which can be computationally intensive). For a single IMU, there are already six separate acceleration and angular velocity axes to be evaluated. In this paper, instead of classifying the signal segments simply as "good" or "bad" quality we propose a learning to rank based approach for the quality assessment of cardiac signals, which is able to determine the relative importance of a signal axis or waveform. We illustrate that the method can generalize between multiple human experts annotated ground truths in automated best axis selection and ranking of signal segments based on their quality.

1. Introduction

Noise in cardiac signals can be originated from various sources, such as body motion or electrical disturbances. Usually, there is a need to perform some kind of pre-processing in order to separate the noisy signal segments (such as motion artifacts) from the rest of the signal. Usually, the signal being monitored contains similar or repeating patterns (such as in the case of ECG signal of a specific lead). In the case of multi-dimensional mechanocardiography (MCG), which is based on joint accelerometers and gyroscopes, both intra and inter-subject variability between persons is large (in comparison with ECG). In addition, the noise varies significantly in both amplitude and frequency, especially if the subject is not still.

When processing ECG signals for example, it has been shown that classifying the signals simply as having "good" or "bad" quality using a binary classifier could provide decrease in the error rates i.e. on the number of false alarms using physiological signal databases [1]. Thus, the detection of noise and motion artifacts forms a crucial pre-processing step in automated diagnosis systems and tele-monitoring [2], since false diagnosis alarms may lead to unnecessary visits to hospitals and cause unnecessary extra burden to the medical staff.

If the signals being measured are classified only as "good" or "bad" quality, a classifier tries to find a separation between the two separate classes. Then, only the good quality signal segments are selected for for further processing. On the other hand, if all the signals are of "good" class/quality, it would still make sense to select only the best ones of these "good" signals for further processing, especially if the signals convey similar information. This could be used to optimize the performance of an automated diagnosis system. One alternative would be to examine the a posteriori probabilities of the classifier outputs and select the largest, but a more preferred option would be to directly use a classifier which has been truly designed to this purpose.

If considering a regression approach, which output is also continuous value - and keeping the application in mind - it might be unnecessary difficult to a human expert to decide which is the exact quality of a given signal segment in some specific continuous scale (such as 1 to 100). Here we propose a pairwise learning to rank approach to solve this issue. Learning to rank can be used to order human preferences (or signals, for instance) based on their relative importance with respect to each other. In the supervised learning scenario proposed, a human expert decides based on two signal segments which one of those more likely is of better quality and which one of lower quality. Thus, the human experts gives an ordering for a signal-

pair shown to the expert. In this paper we concentrate on the quality assessment of cardiac IMU signals. MCG/IMU is a promising non-invasive modality in body sensor network applications due to the wide availability of compact MEMS sensors. Furthermore, our approach can also scale also to other modalities beyond MCG/IMU (such as ECG, PPG, EEG - electroencephalography).

2. Related Work

Cardiac signal quality analysis belongs to a more general class of quality assessment, which has been extensively studied, for example, in image processing [3]. In general, quality assessment can be performed with an existing reference signal, or "blindly" - without any reference. Learning to rank was recently proposed for blind image quality assessment in [4]. The concept of rank based learning has previously been used, for example, to optimizing clickthrough data in the internet search engines [5].

The recent advances of ballistocardiography (BCG) and seismocardiography (SCG) have been reviewed in [6]. Originally, SCG was proposed in [7] to measure the tiny cardiogenic accelerations induced by the heart into the human body. In BCG [8], the subject is monitored through special equipment (such as force sensors) installed e.g. to the weighting scale, chair or the bed where the person is on [6]. Gyrocardiography (GCG) is another non-invasive technique based on chest rotational vibrations induced by the heart movement [9]. BCG, SCG, and GCG have been suggested for various clinical and well-being applications [6, 9].

In the telemonitoring domain, a system based on ECG and PPG was designed [10], which mainly operated in idle mode (to save battery). The aim of the system was heart rate estimation and it was able to detect automatically when a sufficiently good signal segment had been processed. A traditional approach for compensating motion artifacts in ECG is to attach an accelerometer and filtering the ECG [11] through information available from the accelerometer. In [12] adaptive filtering was used for the cancellation of noise in ECG analysis. Motion artifacts have also been studied in the context of a ballistocardiogram measurement of a weighting scale [13].

3. Data Acquisition and Labeling

We collected IMU measurements from 71 healthy volunteers (16 females). The duration of each measurement was approximately 3 min. We used a standard Sony Xperia Z-series smartphone equipped with Google Android OS for the data collection. The smartphone was placed on the chest of the subjects while laying down in a bed. Approximately half of the measurements formed a training set and the others a test set. The captured data (all six accelerometer and gyroscope axes) were pre-processed with FFT (Fast Fourier Transform) brick wall bandpass filter (1 Hz to 45 Hz) to remove both signal bias and noise. Any other noise removal was not performed to the data, thus the data was applied as such after the filtering to the test set and training set generation. Thus, motion artifacts existed in both the test and training sets.

The data was divided into 10 s duration segments, which we consider to be long enough to provide a heart rate estimate or equivalent, but being short enough to provide a human expert a view upon the extent of noise and motion artifacts among an arbitrary pair of signals. Three persons with experience in cardiac signal processing annotated the same 500 pairs of randomly selected image segments according to their relative quality. The human experts were given three alternatives for each of the signal segment-pairs. The alternatives were 1) upper is of better quality, 2) lower is of better quality and 3) skip (if the human expert could not make a decision based on the segments quality). The criteria for quality assessment was that the segment showed better in terms of beat identifiability and was subject to less motion artifacts. A test set was generated in a similar fashion by the same three human experts and it consisted of 200 randomly generated signal pairs for each expert (each expert saw the same segments).

4. Feature Extraction and Learning

We utilize multiscale 1-D local binary pattern (1-D LBP) features [14] to describe the quality characteristics of the IMU signals. We utilized spacings between individual signal samples of 1, 5, 10, 15, 20, 30 and 50 samples to cover both small scale and large temporal scale changes in the signals. The 1-D LBP was calculated by comparing temporal samples within a known spacing to a chosen center sample of a temporal window. If the center sample is smaller or equal to the peripheral sample (within the window), a bit 1 is chosen as the outcome and bit 0 otherwise. All the succeeding bits are concatenated and changed to a decimal, which is between 0 to 255 with eight bits (eight bits used in each spacing in this paper). For instance, a spacing of 5 samples means that the 1-D LBP window is of temporal length of 5*4 samples backwards in time from the center sample and 5*4 samples forward in time from the center sample (in total 41 samples, where there are always 4 non-used samples between two succeeding LBP samples).

We also utilized uniform patterns [15] (LBPs with at most two 1-0 or 0-1 transitions) to reduce the feature vector length. The initial dimension of the feature vector was 256*7 (seven different spacings and for each spacing a separate histogram of length 256), while utilizing the uniform patterns only it became 59*7, i.e. 413 histogram bins. As described, the data was divided into 10
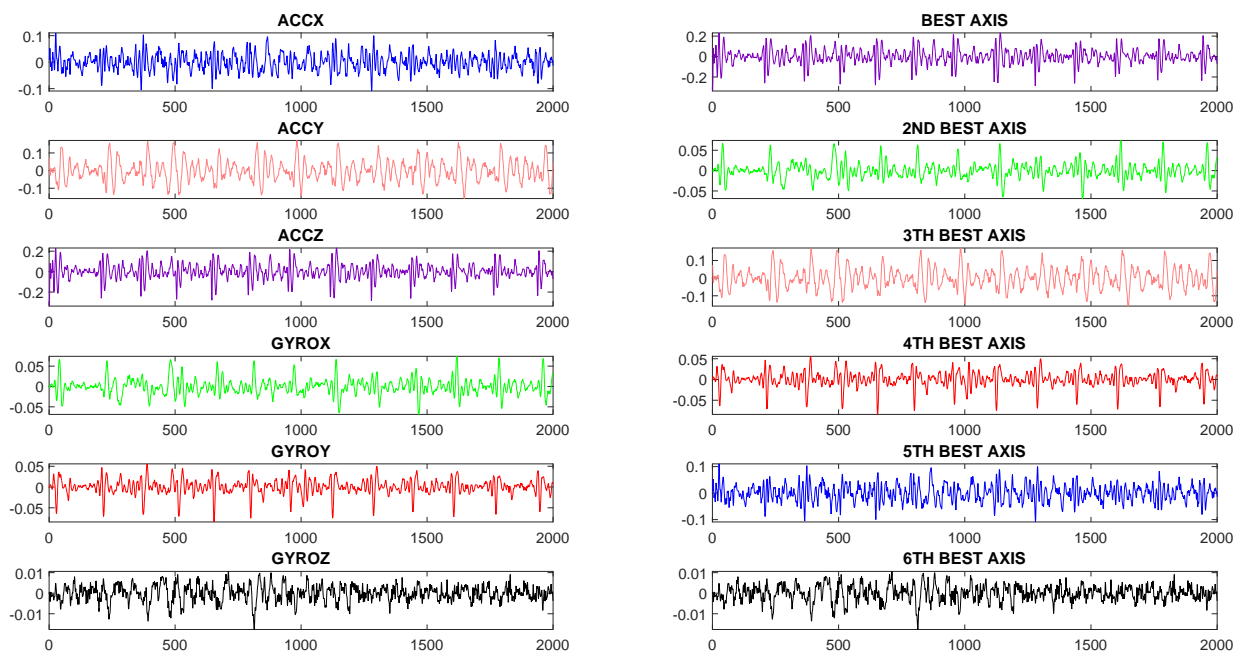
Figure 1. An example of ordering six IMU (3-accelerometer and 3-gyroscope) axes with respect to their relevance/quality. The same axes on the right and on the left are highlighted with the same color. The length of each segment is 10 seconds at 200 Hz sampling rate of the smartphone.

second non-overlapping segments, which were used for the generation of the feature vectors. We used a pairwise ranking support vector machine (SVM) approach available at http://olivier.chapelle.cc/primal/ described at [5] for the learning of the human expert preferences.

Naturally, it may be that a human expert biases the learning results towards his/hers own judgement. If the training generalizes well also to other people, it is probable that the learning has succeeded. We obtained an average rate of 67.8 % in sorting two given signal segments in a correct preference order by their estimated quality/relevance using leave-one-person-out cross-validation (one human expert at a time was used for training and the others were used for testing). With using the same person for both training and testing we obtained an average rate of 73.5 % in the ordering the signal-pairs (without cross-validation of the human experts). In both tests "skipped" signal segment pairs were omitted. Fig. 1, shows an example of all six IMU axes from one of the persons of the test set, and the same signals sorted according to their best relevance. The corresponding axes in the two columns of Fig. 1 are highlighted with the same color.

The tendency that the results are biased towards the opinion of the human expert (i.e. 67.8 % with person-based cross-validation and 73.5 % accuracy otherwise) would not necessarily be a disadvantage. For instance, if instead of signal processing experts, the human experts would be cardiologists, their opinion on sufficient quality might also reflect a characteristics of reliable signal segment for diagnosis purposes. As the ultimate target is to provide efficient tools for automated diagnosis systems based on cardiac data, human involvement could be benefited also this way. Studying whether the method proposed in this paper could be used towards this purpose is, however, left for future work.

5. Discussion

We proposed a method for biomedical signal quality assessment through learning to rank. We utilized six axis IMU signals with 3 accelerometer and 3 gyroscope axes captured with a standard smartphone. In the future we plan to test the developed methods with other modalities such as PPG, ECG and potentionally EEG. Although the focus of this paper was on best axis selection, the learning to rank could be utilized also to more sophisticated purposes, such as improving the performance of automated diagnosis systems utilizing body sensor networks. We utilized 1-D LBP features, which aim at detecting the structure of the signal at various scales. We also tested some other features such as mean and standard deviation of the signal segment, as well as the magnitude and location of the main signal frequency component through FFT, but it seems that they would not improve the classification rate

significantly. We observed also, that increasing the size of the training set generally improved the classification rate, which seems promising, as the size of the training set (500 signal pairs) was still fairly small.

6. Conclusion

We presented a novel machine learning approach for ranking of cardiac signals according to their quality. In our application, the quality of the signal referred to how well the heart beats can be identified within the IMU signal, as well as to the amount of motion artifacts with the given signal segment. In the future, we plan to test the developed algorithm to ECG, PPG and EEG as well as other modalities to aid automated monitoring and diagnosis, with training data labeled by cardiologists. The developed algorithm could part of heart beat extraction process or part of even more advanced automatic diagnosis system, where the selection of a proper signal for further processing performed via learning to rank to save computing resources and improve the detection accuracy.

Acknowledgements

References

[1] Behar J, Oster J, Li Q, Clifford GD. Ecg signal quality during arrhythmia and its application to false alarm reduction. IEEE transactions on biomedical engineering 2013; 60(6):1660–1666.

[2] Redmond SJ, Lovell NH, Basilakis J, Celler BG. Ecg quality measures in telecare monitoring. In Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. IEEE, 2008; 2869–2872.

[3] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 2004;13(4):600–612.

[4] Gao F, Tao D, Gao X, Li X. Learning to rank for blind image quality assessment. IEEE transactions on neural networks and learning systems 2015;26(10):2275–2290.

[5] Joachims T. Optimizing search engines using clickthrough data. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002; 133–142.

[6] Inan O, Migeotte PF, Park KS, Etemadi M, Tavakolian K, Casanella R, Zanetti J, Tank J, Funtova I, Prisk G, Di Rienzo M. Ballistocardiography and seismocardiography: A review of recent advances. Biomedical and Health Informatics IEEE Journal of July 2015;19(4):1414–1427. ISSN 2168-2194.

[7] Bozhenko BS. Seismocardiography– a new method in the study of functional conditions of the heart. Terapevticheskii Arkhiv 1961;33:55–64.

[8] Gordon JW. Certain molar movements of the human body produced by the circulation of the blood. Journal of Anatomy and Physiology 1877;11:533–536.

[9] Jafari Tadi M, Lehtonen E, Saraste A, Tuominen J, Koskinen J, Teräs M, Airaksinen J, Pänkäälä M, Koivisto T. Gyrocardiography: A new non-invasive monitoring method for the assessment of cardiac mechanics and the estimation of hemodynamic variables. Scientific Reports 7 2017; 7(1):6823. ISSN 2045-2322. DOI 10.1038/s41598-017-07248-y.

[10] Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D, Tarassenko L. Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. IEEE journal of biomedical and health informatics 2015;19(3):832–838.

[11] Raya MAD, Sison LG. Adaptive noise cancelling of motion artifact in stress ecg signals using accelerometer. In Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint, volume 2. IEEE, 2002; 1756–1757.

[12] Thakor NV, Zhu YS. Applications of adaptive filtering to ecg analysis: noise cancellation and arrhythmia detection. IEEE transactions on biomedical engineering 1991; 38(8):785–794.

[13] Wiard RM, Inan OT, Argyres B, Etemadi M, Kovacs GT, Giovangrandi L. Automatic detection of motion artifacts in the ballistocardiogram measured on a modified bathroom scale. Medical biological engineering computing 2011; 49(2):213–220.

[14] Chatlani N, Soraghan JJ. Local binary patterns for 1-d signal processing. In Signal Processing Conference, 2010 18th European. IEEE, 2010; 95–99.

[15] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence 2006;28(12):2037–2041.

Address for correspondence:

Olli Lahdenoja
Department of Future Technologies, University of Turku
olanla@utu.fi