Electrocardiogram Quality Classification based on Robust Best Subsets Linear Prediction Error

Kai Noponen, Mari Karsikas, Suvi Tiinanen, Jukka Kortelainen, Heikki Huikuri, Tapio Seppänen

University of Oulu, Oulu, Finland

Abstract

A computationally efficient electrocardiogram (ECG) quality classifier is developed. It is based on the residuals between filtered and observed data, and between the best subset linear predictions without the constant term and the filtered data. Amplitude information is also used. First, the ECG is filtered for essential spectrum band-pass and interference removal. Then, the prediction of each lead is derived only from the information present in three other leads at the same time instant. The prediction coefficients are determined from acceptable quality data using a robust method, and the best lead combinations are found using an exhaustive search.

Trained for maximal accuracy, the classifier achieves 93.2 % accuracy, 96.9 % sensitivity, 80.4 % specificity, 94.5 % positive predictive value, and 88.3 % negative predictive value on training data (positive = acceptable; negative = unacceptable). External blind validation against test data yields an accuracy of 90.0 %.

1. Introduction

Typical problems with ECG acquisition include, among other things, misplaced electrodes, cable switches, poor or no skin-electrode contact, external electro-magnetic interference (EMI), electromyographic (EMG) noise, and artifacts resulting from patient motion [1-3]. Despite the fact that it is important to minimize these problems beforehand with user training and system design, there remains a need to either automatically correct these deficiencies or guide the user in rectifying them [2], as it can improve the quality of the ECG.

Failure tolerant intelligent signal processing techniques can be especially beneficial in ambulatory monitoring applications, and pervasive telemedicine applications in which the user is not a trained professional. The first step towards this direction is to detect problems in the measured signal. Consequently, there is a need for methods to assess the quality of recorded ECG in real-time or near real-time with a low computational complexity which is the aim of the PhysioNet/Computing in Cardiology Challenge 2011 [4,5].

In this paper, we present a computationally efficient ECG quality classifier that is based on quantifying the amount of unwanted signal components such as baseline wander and noise, measuring the agreement of the measured signal to a linear model reconstruction, and exploring the statistics of signal amplitudes.

The rest of the paper is organized as follows. First, the data and the methods are described in Section 2. Then, the results are shown in Section 3. Finally, the results are discussed in Section 4.

2. Methods and materials

We use a two-step feature extraction process to describe the quality of the ECG as illustrated in Figure 1 at a high abstraction level. First, the signal is filtered for band-pass and power-line interference removal. The residual between the original signal and the filtered signal varies together with many quality decreasing components such as baseline wander, EMI, EMG-noise, and artifacts.

Second, the filtered signal in each lead is predicted linearly using only three other leads. The residual between the filtered signal and the predicted signal shows the agreement between the model and the observation. Due to the dipolar nature of the heart as a source of electric field [6], the agreement is expected to be good when the signal originates from the heart and contains little in-band interference that passes the filtering stage.

Third, the sizes of the residuals are measured with block-wise L1-norm. Finally, the norms and signal range information is used to classify the quality of a recording as either acceptable or unacceptable. The steps are described in more detail in the following sub-sections.

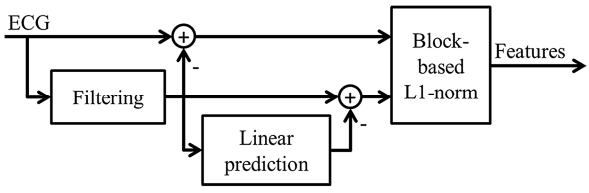

Figure 1. A high-level flow diagram of the proposed method.

2.1. Data

The Set-A of the PhysioNet/Computing in Cardiology Challenge 2011 [4,5] is used for training. It contains 1 000 recordings with quality labelled for reference either as acceptable (773 of 1 000) or unacceptable (225 of 1 000). Instances without a label are considered to be of indeterminate quality (2 of 1 000). For our intents and purposes, the indeterminate records are excluded from the learning dataset.

In testing, we use the corresponding Set-B of the challenge. It contains 500 ECG recordings but without the reference quality information. This partitioning of the data into the two sets, A and B, provides a hold-out validation for the results. Both of these datasets contain standard 12-lead ECG recordings with 10 second duration acquired with conventional ECG machines. Each lead is sampled at 500 Hz with 16-bit resolution, and is guaranteed to include bandwidth from 0.05 Hz to 100 Hz.

2.2. Filtering

In this step, we use the characteristic frequency bandwidth of ECG to remove unwanted signal sources that may have been mixed into the measurement. This enables us to assess the strength of these components with the assumption that this gives information on the quality of the ECG. The filtered ECG is also used in the linear model described in the next section.

The uses of ECG are various, ranging anywhere from simple rhythm monitoring to detailed morphological analysis. Consequently, there exists a great deal of diversity of filtering requirements in the literature. Although the bulk of the power spectrum is narrower, typical recommendations imply a low frequency cutoff of 0.67 Hz at most, and a high-frequency cutoff of 150 Hz at least together with phase response considerations [2].

In accordance with these specifications, we use a fourth order Butterworth high-pass filter with 0.6 Hz cutoff frequency. The filter reduces baseline wander, and the effects of respiration and movement. Similarly, a fourth order Butterworth low-pass filter with 150 Hz cutoff is used to counteract higher frequency noise such as EMI and EMG. Power-line related interferences that fall within these frequencies are handled using notch filters at 50 Hz, 60 Hz, 100 Hz, and 120 Hz.

All the filters are infinite impulse response (IIR) type filters with non-linear phase response. In order to obtain zero phase distortion, the filters are applied bi-directionally to each lead of the ECG recording. This addition of the time-reverse processing doubles the filter order as well as the computational cost compared to normal causal one-way application used in on-line applications. Due to the small number of filter coefficients, the computational cost is low nevertheless.

2.3. Linear prediction

In this step, ECG is modelled using multiple linear models without the constant term. Only the Set-A records with acceptable reference quality label are used in coefficient estimation.

To utilize the inherent dipolarity of the heart and to prevent unnecessary crosstalk, the signal in each lead is predicted using only three other leads, a small disjoint subset of the whole lead set. It should be noted that for each lead, the selected subset can be different. As a further restriction, only one limb or augmented lead is allowed to be selected into each of the subsets used for prediction because of the direct mathematical interdependences between these leads.

For each lead, the best subset is found using an exhaustive search through all the possible combinations. Since the learning data is expected to contain outliers, noise, and other non-ECG related phenomena, the coefficients are learned using robust multivariate regression by iteratively reweighted least squares (IRLS) method with a bisquare weighting function [7].

Noisy or otherwise bad quality signal segments in a lead used to predict another one directly affect the quality of the prediction. Consequently, two multivariate models are constructed to counteract this leaking phenomenon. The second model is constructed with the additional restriction: For each lead, only those other leads that have not been selected to the first model can be utilized. Predictions are made using both the models concurrently.

2.4. Block-based norm calculation

In this step, the sizes of the residuals after filtering and prediction are estimated. First, the signal in each lead is divided into disjoint blocks of constant duration. This division to blocks enables locating discrepancies in time.

Second, the norm of the residual is calculated block-wise. Since the aim is to describe the size of the residual as a whole, the L1-norm is chosen for robustness against outliers within the block. The norm evaluation is performed for the residual between filtered and original data, as well as the residuals between both the predictions and the filtered data.

Finally, we choose the smaller one of the prediction residual norms for each block. The reasoning behind this is that the larger one is usually due to the cross-talk phenomena described in the previous section. It should also be noted that normally the performance of the second model is worse than the first model by construction. Consequently, a re-evaluation of the situation is justified when the second model suggests a smaller sized residual.

By varying the block-length, time resolution can be traded to statistical stability: The shorter the block duration, the higher the variance of the residual norm.

2.5. Classification

A simplified classifier with only a few parameters is used to avoid problems in over-training and the curse of the dimensionality. What is more, the simplistic approach is chosen to show the potential of the selected features to describe the quality of ECG. Our classification scheme utilises some heuristic knowledge about ECG signals. It is based on the sizes of the L1-norms of the filtering residual and the linear prediction model residual. In addition, the amplitude variation is taken into account.

More precisely, the quality of an ECG record is classified as acceptable if and only if all the following conditions are fulfilled; otherwise it is unacceptable:

- The signal range (max-min) in none of the leads is under the global disconnection threshold *D*.
- The signal amplitude is within the lead-wise lower and upper limits *Li* and *Ui* for at least *Q* percent of the total duration in at least *N* leads.
- For at least *M1* limb/augmented leads and at least *M2* chest leads, the sum of the filtering L1-norm residual blocks is below the threshold *Fi* which is separate for the two subsets of leads.
- For at least *O1* limb/augmented leads and at least *O2* chest leads, the sum of the prediction L1-norm residual blocks is below the threshold *Gi* which is separate for the two subsets of leads.

To reduce the amount of parameters, same thresholds are applied to several channels. However, as a heuristic rule, limb and augmented leads are treated separately from the chest leads due to signal scale considerations.

The residual is not scale invariant, that is, small-amplitude signals lead to small-amplitude residuals. This can be problematic with disconnected or poorly connected electrodes commonly encountered. Therefore, the signal range is also included in the decision.

All the parameters are learned from the Set-A data using an exhaustive grid search. The lower and upper limits, *Li* and *Ui*, for the signal amplitude are selected as the *p*:th and *100-p*:th percentile of the training data labeled acceptable. Although the limits are individual for each lead, they are controlled by the single parameter *p*.

3. Results

3.1. Linear prediction models

Table 1 shows the linear prediction coefficients learned from the filtered Set-A data with acceptable reference quality label. Each row of the table describes how the prediction of the lead in the first column is formed as a linear combination of the leads in the other columns together with their weights. Similarly, the coefficients for the second multivariate model are enumerated in Table 2.

Table 1. Coefficients for the primary prediction model.

Pred. lead	Lead: coef 1	Lead: coef 2	Lead: coef 3
I	aVL: 0.8747	V1: -0.2046	V4: 0.2543
II	aVF: 0.8639	V1: -0.0859	V6: 0.3507
III	aVL: -1.1196	V1: -0.1980	V4: 0.2454
aVR	I: -0.6745	V2: 0.0594	V4: -0.2123
aVF	III: 0.7528	V1: -0.1959	V4: 0.1922
aVL	III: -0.7556	V1: -0.1963	V4: 0.1930
V1	I: -0.4896	V2: 0.4232	V6: -0.1402
V2	III: -0.2943	V1: 0.6341	V3: 0.5927
V3	V2: 0.2776	V4: 0.9306	V6: -0.2412
V4	V1: -0.0105	V3: 0.4238	V5: 0.6653
V5	V3: -0.0786	V4: 0.5299	V6: 0.5864
V6	aVR: -0.1788	V1: -0.1186	V5: 0.6434

Table 2. Coefficients for the secondary prediction model.

Pred. lead	Lead: coef 1	Lead: coef 2	Lead: coef 3
I	aVR: -1.0364	V2: 0.0671	V3: -0.1051
II	aVR: -0.9829	V2: -0.0663	V3: 0.1063
III	aVF: 1.1536	V5: -0.0641	V6: -0.3143
aVR	II: -0.5800	V1: 0.0874	V6: -0.2252
aVF	II: 0.8894	V3: 0.0045	V6: -0.2594
aVL	I: 0.9066	V3: -0.0558	V6: -0.2923
V1	aVR: 0.6968	V3: 0.3754	V5: -0.1285
V2	II: -0.2263	V4: 0.9449	V6: -0.5634
V3	II: 0.2268	V1: 0.6728	V5: 0.8266
V4	aVF: 0.1948	V2: 0.3259	V6: 0.8627
V5	II: 0.7880	V1: -0.2413	V2: 0.2123
V6	I: 0.3215	V3: -0.2591	V4: 0.6975

3.2. Classifier performance metrics

Training to maximal accuracy with the Set-A and the 250 sample (500 ms) block length, the presented method achieves 93.2 % accuracy, 96.9 % sensitivity, 80.4 % specificity, 94.5 % positive predictive value, and 88.3 % negative predictive value on the training set itself. It should be noted that here a positive is used to denote acceptable quality, and a negative unacceptable quality. In addition, according to the PhysioNet/Computing in Cardiology Challenge 2011 submission system for the Event-1 [5], the blind validation accuracy against the Set-B is 90.0 %.

For reference, two research scientist with experience from ECG signal processing manually annotated records from the separate halves of the Set-A as either acceptable or unacceptable in quality in a non-blind manner. In other words, a preliminary reference classification but not the final one was already published at [5]. Holding the final reference classification as the truth information, their labeling accuracies are 89.4 % and 82.6 %.

4. Discussion

In this study, we present a computationally efficient but robust approach to ECG quality classification. Our results indicate that the accuracy reaches or surpasses the accuracy of individual annotators with some experience in the field. Based on the final scores (89.6 % - 93.2 %) of the ten best performing entries to the Event-1 of the Physionet Challenge 2011 [5], we may conclude that our accuracy results are in line with what can be expected. Moreover, the controlled decrease of accuracy in blind validation shows the ability to generalize.

A key advantage of the presented method is that no explicit assumptions are made about the actual shape of the signal which can exhibit large intersubject and intrasubject variability. For instance, the former can occur due to natural variability in the body size and shape as well as differences in health. The latter can occur due to changes in the electrolyte balance, onsets of diseases, or effects of medications, to name a few.

Noise filtering and signal amplitude type quality measures used in this work are well-known, see e.g. [8]. A major shortcoming of filter usage alone is the inability to handle in-band noise and artifacts. Often, this leads to more elaborate processing of the signal morphology, cf. [9]. In contrast, our method requires no R-peak detection, delineation, alignment, or any higher level ECG processing tasks at all. Consequently, it is not dependent on the success of such tasks. The expected expression of the signal is intrinsically controlled through the linear predictive model that captures the dipolar nature of the heart. It captures knowledge about the timing of events between leads and also how the shapes of the signals in different leads are expected to match to each other. We argue that this kind of information on the dependencies between the leads is often used by domain experts in the inspection of ECG recordings.

The computational complexity of the learning phase is significant due to the exhaustive search and robust regression methods. More importantly, however, the method is very efficient when classifying new ECG recordings. As a result, it can be implemented in devices such as mobile phones or wearable ECG monitoring devices. Since our classifier is simplistic by design, the presented feature extraction steps offer potential to reach even better results when using more elaborate classifiers, more learning data, and more rigorous cross-validation.

To conclude, we would like to emphasize that the quality of ECG is a multifaceted concept difficult to capture as a single number or as a clear-cut classification. In one point of view, it can be argued that the accuracy and reliability of an end result, be it the rhythm or a diagnosis, determines if the ECG acquisition was of acceptable quality. In another point of view, the quality of ECG is largely separated from the diagnostic information contained in it and relates to how well the measured signals represent the electric activity of the heart and nothing else. Our approach has been designed to yield some insight to both of these notions of quality with a bias to the latter. According to the results, it has been fruitful and shows potential for further development.

Acknowledgements

This work was supported by the Walter Ahlström Foundation, and the BraveHealth EU-project.

References

[1] García-Niebla, J, Llontop-García P, Valle-Racero JI, Serra-Autonell G, Batchvarov VN, De Luna, AB. Technical Mistakes during the Acquisition of the Electrocardiogram. Ann Noninvasive Electrocardiol 2009;14(4):389-403.

[2] Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, et al. Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part I: The Electrocardiogram and Its Technology A Scientific Statement From the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology. J Am Coll Cardiol 2007; 49:1109-27.

[3] Pahlm O, Hammill SC. Quality improvement in electrocardiogram recording and interpretation [editorial]. J Electrocard 2008;41(5):367-9.

[4] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circ J 2000;101(23): e215-20.

[5] PhysioNet/Computing in Cardiology Challenge 2011 [Internet; Updated 2011 Aug 17; cited 2011 Sep 8]. Available from: http://physionet.org/challenge/2011/.

[6] Malmivuo J, Plonsey R. Bioelectromagnetism — Principles and Applications of Bioelectric and Biomagnetic Fields. New York: Oxford University Press, 1995.

[7] John F. Applied regression analysis, linear models, and related models. Thousand Oaks (CA): Sage, 1997:405-11.

[8] Redmond SJ, Lovell NH, Basilakis J, Celler BG. ECG quality measures in telecare monitoring. In: Conf Proc IEEE Eng Med Biol Soc 2008. Vancouver (Canada): IEEE, 2008:2869-72.

[9] Vaglio M, Isola L, Gates G, Badilini F. Use of ECG Quality Metrics in Clinical Trials. Computing in Cardiology 2010;37:505-8.

Address for correspondence.

Kai Noponen
University of Oulu
Department of Computer Science and Engineering
PO Box 4500
FI-90014 UNIVERSITY OF OULU
Finland