Two Simple Multi-lead ECG Quality Assessment Algorithms

Daizong Yang*, Yue Zhang
Laboratory of Embedded Systems and Technology
Graduate School at Shenzhen, Tsinghua University
1850108730@qq.com

Abstract. The purpose of ECG signal quality assessment is to eliminate the poor-quality ECG signal, to increase the accuracy of ECG feature point detection and reduce the misjudgements of cardiac arrhythmia. At present, the main problem of ECG quality assessment is that it adopts R-wave related features. But what we hope is that the R wave shouldn’t be detected until the signal quality is determined. We present two ECG quality assessment schemes based on AR model and SVD decomposition from different perspectives. The SVD method has higher sensitivity and accuracy. In practice, we can choose whether to use it alone or to fuse two methods according to different needs.

1. Introduction
Although the R-peak detection algorithm in our previous article [1] has a high overall accuracy, miss or false detection mostly focus on a few records. The error rates of record 105, 108 and 203 in MIT-BIH database [2] are all over 1%. These errors are mainly due to the poor signal quality of these records, especially serious EMG noise or baseline drift interference. The inaccurate segmentation of the heartbeat will lead to the misjudgements of the arrhythmia analysis by extracting inappropriate features from the heartbeat. Fig. 1 shows a segment with poor signal quality. Although most of the beats are normal type (N) heartbeats, their shapes differ greatly from those of good quality signals. Our R-wave detection algorithm mainly produces error detection on such ECG fragments.

Figure 1. ECG Fragment of record 105 in MIT-BIH database

Figure 2. ECG signals collected during intense exercise

In addition, the most vulnerable disturbances to dynamic ECG signals are large baseline drift and sudden abnormal increase of signal amplitude. Fig. 2 shows the signals collected by using our own patch-type ECG acquisition device [3] during intense exercise.

There are many reasons for the poor signal quality, which are the result of EMG noise, power frequency noise, baseline, high frequency noise and so on. For real-time dynamic ECG data, it may also be caused by poor contact with electrodes due to skin contraction in the chest during exercise. Some classical algorithms have tried to filter ECG signals, such as wavelet transform, Kalman filter...
[4], PCA or ICA-based methods [5], template-based adaptive filtering [6], etc. But these methods are suitable for situations where the signal is not seriously disturbed by noise or only for certain noise, such as [6] only for white or pink noise, [5] is more suitable for artifacts.

Another point of view is that we remove signals whose waveform is severely damaged. ECG signals are collected 24 hours a day in real time by our patch-type ECG acquisition equipment and signal quality performs well most of the day. So removing part of the signal will not affect our diagnosis of a patient with a certain disease, while reducing the storage and computing costs of the signal acquisition equipment. It provides better signal for the diagnosis of arrhythmia and improves the accuracy.

A concise method is adopted in [7]. Firstly, it judges whether the signal has lead off, and then uses the number of R peaks to measure the high frequency noise. When the number of R peaks in a certain lead is less than half of the total number, it judges that there is noise. Detecting the starting point of R wave, determining the PQ point, calculating the RMS of the signal near the PQ point, and judging the serious low frequency noise when RMS exceeds a certain threshold value. In [8], the vector graph reconstructed from 12-lead ECG is used for evaluation, but the reconstruction matrix algorithm is complex and computational complexity is large. Four criteria are defined in [9] to measure the quality of ECG signal. Pulse detection criterion measures whether a signal has a high pulse waveform, the lead crossover points criterion measures whether the baseline drift is serious, the empty lead criterion judges whether there is ECG waveform, and the QRS detection robustness criterion tests the consistency of the number of R wave in each lead. In [10], four metrics are also defined to detect whether the electrodes are in good contact, whether there is Gauss noise, whether there are pulses and the deviation of R-peak detection between different leads. The high-frequency band energy, low-frequency band energy and effective energy of ECG signal are estimated by short-time Fourier transform in [11], and then the R-wave amplitude obtained by QRS detection algorithm is added as feature, and the cascaded single-condition decision rules are used for classification. In [12], the existence of lead off is judged, and then the heart rate characteristics and power spectrum distribution characteristics obtained by R-wave detection are fused to fed into Neural network.

However, the most serious problem with these methods is that most of them use R-wave-related features. We hope that ECG feature points should be extracted after eliminating ECG fragments with poor signal quality, so as to avoid additional computational overhead.

We evaluate the ECG quality from two different perspectives. One is based on the linear model of time series, which is used to cross-validate the time series, to find out whether there are abnormal waveforms in the sequence. Another is to use singular value decomposition to analyse the proportion of the principal components of 12-lead signals. The distribution of the principal components of normal and abnormal signals is different.

2. ECG Quality Assessment Based on AR Model

2.1. Model selection

AR (Autoregressive Model), MA (Moving Average Model) and ARMA (Autoregressive Moving Average Model) are common methods for stationary time series modelling [13].

The formula of AR(p) model is as follows:

\[ X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t, t \in \mathbb{Z} \]

(1)

The characteristic polynomial is:

\[ A(z) = 1 - \sum_{j=1}^{p} a_j z^j \]

(2)

In the above formula, p is the order of AR model, \( \varepsilon_t \sim WN(0, \sigma^2) \) is white noise sequence. If the coefficient \( a_j \) makes the characteristic polynomial have no zero roots on the unit circle, the generated
time series $X_t$ is a stationary sequence, if there is zero root on the unit circle, it is a non-stationary sequence, and if there is zero root inside the unit circle, the sequence is non-causal.

The most commonly used model in practice is AR model. Because the coefficient $a_j$ in AR model is easy to estimate, the simplest way is to solve Yule-Walker equation which is composed of autocorrelation coefficients of time series $X_t$. The Yule-Walker equation is a system of linear equations, which can be solved by Levinson-Durbin recursive algorithm when the order $p$ is large [13]. If the ARMA model is adopted, not only the order $p$ of autoregression but also the order of sliding average of white noise needs to be estimated. Moreover, when the order of AR model is high enough, it provides a reasonable approximation for ARMA model. AR model is also often used in linear prediction of speech signals.

2.2. Order determination of AR model

In order to build AR model of ECG signal, it is necessary to check whether the ECG signal with good quality conforms to this model and determine the order $p$ of the model, then estimate the coefficients $a_j$ and $\sigma^2$.

Theoretically, according to the Cramer’s decomposition theorem of time series, any time series can be decomposed into deterministic trend part determined by polynomials and non-deterministic random sequence composed of white noise. AR model can fit the uncertain random sequence part, and the deterministic part of polynomial is also called baseline drift in ECG signal. Therefore, ECG signal with high quality can be regarded as a stationary sequence consisting of linear combination of white noise after eliminating baseline drift.

To determine whether AR fitting can be used, we use augmented unit root test (ADF, Augmented Dickey-Fuller) [14] here and determine the upper limit of order $p$ in the test. AR($p$) model can be rewritten as the following formulas (3-5), where $\nabla$ is a difference operator:

$$\nabla X_t = \gamma X_{t-1} + \sum_{j=1}^{p-1} \varphi_j \nabla X_{t-j} + \epsilon_t$$

$$\gamma = \sum_{j=1}^{p} a_j - 1$$

$$\varphi_j = \sum_{i=j+1}^{p} a_i$$

The zero hypothesis of ADF test is $H_0: \gamma = 0$, that is, if the sequence is modeled by AR model, the characteristic equation will have unit roots, and AR model is a non-stable system. The alternative assumption is $H_1: \gamma < 0$, and the AR model is stable.

In the test, we first determine that the empirical upper limit of $p$ is $p_{max} = [12 \cdot (N/100)^{1/4}]$, where $N$ is the number of sequential samples. If the P value of the t statistic of $\varphi_{p-1}$ is less than 0.05, the parameters are considered significant. Then the ADF test is carried out to determine whether the sequence can be modeled by AR. If the significance level of the parameter $\varphi_{p-1}$ is greater than 0.05, the procedure mentioned above is repeated after reducing the order $p$. In this way, the order $p$, which rejects the zero hypothesis and the highest order coefficients of the model are significant, is regarded as the ultimate upper-bound order of AR modelling, and is denoted as $p_{limit}$.

The above steps only determine the range of order, but we also need to determine the best AR fitting order. If only a single ECG data is processed, the autocorrelation (acf) and partial autocorrelation (pacf) coefficients of the observed sequence are commonly used methods, and they are only qualitative methods. The commonly used quantitative method is Bayesian Information Criterion (BIC).

$$BIC(p) = \ln \sigma^2 + \frac{(p+1) \ln N}{N}$$
\[ \hat{\sigma}^2 \] is the estimated variance of white noise, \( N \) is the sample number of time series, \( p \) is the AR order. The integer \( p \) which minimizes the BIC value is the best order of AR model in the interval \([1, p_{\text{limit}}]\).

The best linear one-step prediction of AR\([p]\) is:

\[
\hat{X}_{t+1} = \sum_{j=1}^{p} a_j X_{t+1-j}
\] (7)

When \( \epsilon_t \) is white Gaussian noise, the linear optimal prediction becomes the best prediction in the sense of minimum mean square error [13].

2.3. Coefficient of determination (goodness of fit)

The coefficient of determination is used to measure the goodness of fit between AR linear prediction and the original signal. The coefficient reflects the proportion of the variance of dependent variable that can be explained by the variance of independent variable. The coefficient of determination \( R^2 \) is defined as follows, where \( SS_{\text{res}} \) is the sum of residual squares and \( SS_{\text{tot}} \) is the sum of total deviation squares:

\[
R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}
\] (8)

\[
SS_{\text{res}} = \sum_{i=1}^{N} (X_i - \hat{X}_i)^2
\] (9)

\[
SS_{\text{tot}} = \sum_{i=1}^{N} (X_i - \bar{X})^2
\] (10)

3. ECG quality assessment based on singular value decomposition

SVD is often used as a method to calculate matrix eigenvalues and eigenvectors in principal component analysis (PCA) and blind source separation (BSS). PCA and ICA can be used in signal filtering [15], which are common means of processing multivariate time series. Both can decompose multivariate time series into different components. The difference is that PCA can also distinguish the proportion of different components in the original signal by the size of eigenvalues. BSS method has two major uncertainties [16].

\[
\hat{S} = DPS
\] (11)

In (11), \( S \) is the source signal vector, while \( \hat{S} \) is the estimation of the source signal obtained by BSS, \( D \) is the diagonal matrix (the diagonal elements are not necessarily the same), \( P \) is the permutation matrix. So, we can't get the order of the source signal, and the estimated signal is multiplied by different scale factors, so we can't estimate the proportion of each source component. In practice, the source components obtained from BSS separation are normalized to a norm of 1. In ECG quality assessment, even if we separate the source signal with noise, we cannot know whether the noise component is large enough to affect the quality. Thus, the process of source separation is not necessary, but the commonly used BSS processing method, whitening, can give us inspiration.

There are two functions of whitening, one is to determine the number of source signals, the other is to make the zero-delay covariance matrix of whitened signals a unit matrix. In [17] [18], the robust whitening method is not to use SVD for a single covariance matrix, but to make up \( m \times m \) matrix of several covariance matrices corresponding to different delays, where \( m \) is the number of observed variables.

\[
L = [M(\tau_1) \ldots M(\tau_J)]
\] (12)

where \( \tau_i \) is the delay and \( M(\tau_i) \) is the covariance matrix. After SVD, the rank of \( L \) is determined according to the eigenvalue matrix, which is the number of sources. Please refer to [19] for more details of the singular value decomposition theorem and how to determine the validity of eigenvalues.
In the evaluation of ECG quality, we divide the observed multi-lead signals into frames, which is a common pre-processing method for speech signals and is mainly suitable for non-stationary signal. We assume that waveform in the same frame is relatively stable, so that the frames including abnormal waveform can be distinguished from the normal frame, and if necessary, a window function can be added to each frame to reduce the correlation of different frames. The zero-delay covariance matrix is calculated for each frame, and a new covariance matrix is constructed in the form of (12). Then the eigenvalues obtained by SVD are directly used as features for quality evaluation.

4. Result and Discussion

4.1. Database
The data we used came from Physionet’s ECG Quality Assessment Challenge Database [20]. Each record in this database is twelve-lead (I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6) data collected by mobile phones in real time for 10 seconds. There are total 998 records, which are labeled as ‘accept’ or ‘unaccept’. Sampling frequency is 500 Hz and resolution are 16 bits.

The ECG in the database was collected by nurses, technicians and volunteers trained to varying degrees. In practice, ECG recorders are not necessarily experienced. Since the goal of the database is to investigate whether software can help laymen to collect high-quality electrocardiograms reliably, the collected data include electrocardiograms collected by volunteers with minimal training.

ECG is graded and scored by a group of people with different professional knowledge in ECG analysis. Each data is independently scored by 3 to 18 people. Data with an average score of more than 0.7 is classified as ‘accept’, otherwise as ‘unaccept’.

4.2. Support vector machines for binary classification
We adopt a two-class SVM with relaxation variables and RBF kernel $\exp\left( -\frac{\|x_i-x_j\|^2}{\sigma^2} \right)$.

$$\min \frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^{m} \xi_i$$
$$\text{st. } y_i(\omega^T \phi(x_i) + b) \geq 1 - \xi_i \quad (13)$$
$$\xi_i \geq 0$$

Two-thirds of the two types of samples in the database are taken as training set, and the remaining are taken as testing set. The training set are stratified sampling to ensure that the original proportion is maintained in each fold. The 10-fold cross validation is used to determine the hyperparameters C and $\sigma$.

4.3. Optimum order and energy proportion
$p_{\text{limit}}$ is determined by ADF test in modeling ECG sequence by AR. Fig. 3 shows that the order p of each lead is no more than 15, and we make $p_{\text{limit}} = 15$.

Secondly, we search the integer values in the $[1, p_{\text{limit}}]$, determine the best order of each data according to BIC, and make a statistical histogram. In Fig. 4, we take lead I as an example. At least 90% of the models have the best order less than 10, and other leads have similar situation. In subsequent experiments, we fix the order $p = 10$ when fitting each data.
The length of each ECG record is 10s and the sampling frequency is 500Hz. We divide each lead into 10 non-overlapping frames, which basically contains a complete heartbeat in each frame. We use the previous frame to fit AR model and use the model to make the best one-step prediction for the data of the latter frame. In fact, this is equivalent to recursive filtering for the data. The closer $R^2$ is to 1, the better the fitting degree is. This shows that there is no variation between the frame, and the closer $R^2$ is to 0, indicating that there is variation between the two frames.

The fitting result of an ‘unaccept’ type data is shown in Fig. 5. The fitting residual of abnormal beats is obviously larger than that of normal beats. We average the nine determinant coefficients of each lead as a feature of the lead.

The average determinant coefficient approaching 1 can only indicate that there is no obvious sudden change in the lead. One extreme case is that the lead signal is a horizontal straight line (only the DC level). The fitting degree is very good, but there is no effective ECG component. Another example is shown in Fig. 6. The large square pulse amplitude in the lead obscures the effective component of the ECG waveform. When we amplify the signal near the square pulse, we will find the perfect ECG waveform. All of these suggest that we need a feature to measure the proportion of effective ECG component in the whole lead.

Figure 3. The upper limit of AR order of 12-lead ECG

Figure 4. Statistical Histogram of Optimum Order for Lead I

Figure 5. AR fitting of ECG lead with abnormal beats

Figure 6. "Unaccept" signal with square pulse and detail amplification signal
According to [21], the energy of normal ECG waveform is mainly concentrated in 1-40 Hz, and the sampling frequency of data is 500 Hz. We use the average periodogram method (Welch) [22] to calculate the power density spectrum of 10s ECG data, and then the ratio of the energy of the signal in the frequency range of 1 ~ 40Hz to that in the frequency range of 0 ~ 250Hz is calculated. We multiply the energy ratio by the average determinant coefficient as the final feature of the lead, then each record containing 12-lead data generates 12-dimensional feature vector. The closer the feature of each dimension is to 1, the better the quality of the lead is.

4.4. Comparison and Result
The signal is divided into 10 frames when using singular value decomposition (SVD) method for quality evaluation. The zero-delay covariance matrix of 12-lead data in each frame is calculated. The ten 12×12 square matrices are constructed as a new covariance matrix in the form of (12). The 12-dimensional eigenvalue vector is obtained by SVD as the feature.

We use sensitivity (Se), specificity (Sp) and accuracy (Acc) to evaluate the algorithm. Sensitivity represents the proportion of the real good-quality signals detected to all acceptable signals; specificity represents the proportion of the poor-quality signals to all unacceptable signals; and accuracy represents the proportion of all correctly detected signals to all samples. The formulas for calculating each index are as follows:

\[
Se = \frac{TP}{TP + FN} \\
Sp = \frac{TN}{TN + FP} \\
Acc = \frac{TN + TP}{TN + FP + TP + FN}
\]

Table 1 compares the results of the proposed algorithm with those of other algorithms.

| Methods                | Sensitivity | Specificity | Accuracy |
|------------------------|-------------|-------------|----------|
| Arie C Maan [8]        | 97.0%       | 75.1%       | 92.2%    |
| Hayn D [9]             | 96.1%       | 84.0%       | 93.4%    |
| Liu C [10]             | 90.67%      | 89.78%      | 90.0%    |
| Johannesen [7]         | 91.0%       | 85.0%       | 88.0%    |
| Di Marco L Y [16]      | —           | —           | 92.77%   |
| Zhang Y [12]           | 95.0%       | 86.7%       | 93.1%    |
| AR                     | 95.0%       | 82.7%       | 91.6%    |
| SVD                    | 96.5%       | 88.0%       | 94.2%    |
| AR+SVD                 | 93.5%       | 93.3%       | 93.45%   |

The results in Table 1 show that the method based on AR model alone is not excellent in terms of sensitivity and specificity, while the SVD method achieves high sensitivity and accuracy. A simple fusion decision-making scheme is that the quality of ECG signals is evaluated independently based on AR and SVD. If and only if the results obtained by the two methods are ‘accept’, the signal is judged to be of good quality. If the result of any method is ‘unaccept’, the signal is judged to be of poor quality.

As shown in Fig. 7, there is only one abnormal beat in each lead, but the amplitude of this beat does not increase significantly compared with the normal part. Therefore, the principal components of
this record are not significantly different from those of the good quality signal, so the distribution of eigenvalues obtained by SVD is not abnormal. However, AR method can recognize this abnormal beat (see Fig. 5), so AR method is an effective complement to SVD. The combination of AR and SVD method can improve the recognition rate of poor-quality signals, but at the cost of reducing sensitivity.

Therefore, in practical applications, if you want to collect as many good quality signals as possible, you can use SVD alone. If you want to strictly control the quality of the collected signals and eliminate the poor-quality signals as far as possible, you can use the method of AR and SVD fusion.

Figure 7. ECG waveform suitable for fusion method

Finally, I would like to point out a mistake in [12], where it is believed that the record is classified as ‘unaccept’ as long as there is leadoff (i.e., the signal has only DC level). To illustrate this error, we use the signal numbered 253789 as an example, which is shown in Fig. 8. Although the record is acceptable, the lead is dropped.

Figure 8. 12-lead ECG waveform No. 253789

5. Conclusion
Firstly, the reason of ECG quality assessment is to reduce the misjudgements of arrhythmia and the calculation and storage overhead of real-time monitoring system. Then, the research progress of ECG quality assessment is introduced. The main problem is that most of the algorithms adopt the characteristics related to R-wave detection. We hope that the feature points of the signal will not be detected in the process of quality assessment, and the follow-up processing will be carried out only when the signal is judged to be of good quality. We present two different ECG quality assessment schemes from different perspectives. The SVD method has higher sensitivity and accuracy. In practice, we can choose whether to use it alone or to fuse it with AR according to different needs.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (No.61571628)
References

[1] Yang, Daizong and Zhang, Yue, “A real-time QRS detector based on low-pass differentiator and hilbert transform,” 2018 International Forum on Construction, Aviation and Environmental Engineering-Internet of Things. v 175, July 2, 2018

[2] Moody, G. B., and R. G. Mark. "The impact of the MIT-BIH arrhythmia database." IEEE Engineering in Medicine & Biology Magazine 20.3(2002):45-50

[3] Yang, Daizong and Zhang, Yue, "Wireless Wearable Patch-Type ECG Acquisition Device for Long-Term Ambulatory Monitoring," 2018 International Conference on Mechatronics Engineering and Computer Sciences. Shenyang, China, pp. 1369–1372, May 2018.

[4] Sayadi, Omid, and M. B. Shamsollahi. "ECG Denoising and Compression Using a Modified Extended Kalman Filter Structure." IEEE transactions on bio-medical engineering 55.9(2008):2240-2248.

[5] Romero, I. "PCA and ICA applied to noise reduction in multi-lead ECG." Computing in Cardiology IEEE, 2011.

[6] Yan, Jingyu, et al. "Self-adaptive model-based ECG denoising using features extracted by mean shift algorithm." Biomedical Signal Processing and Control 5.2(2010):103-113.

[7] Johannesen L. “Assessment of ECG quality on an Android platform,” Computing in Cardiology. IEEE, 2011:433-436

[8] Maan, A. C, et al. "Assessment of signal quality and electrode placement in ECGs using a reconstruction matrix." Computing in Cardiology IEEE, 2011.

[9] Hayn, Dieter, B. Jammerbund, and Schreier, Günter. “QRS detection based ECG quality assessment." Physiological Measurement 33.9(2012):1449-1461.

[10] Liu, Chengyu, et al. "Real-time signal quality assessment for ECGs collected using mobile phones." Computing in Cardiology IEEE, 2011.

[11] Di Marco, Luigi Yuri, et al. "Evaluation of an algorithm based on single-condition decision rules for binary classification of 12-lead ambulatory ECG recording quality." Physiological Measurement 33.9(2012):1435-1448.

[12] Zhang, Yue, and Z. Hou. "An algorithm for evaluating the ECG signal quality in 12 lead ECG monitoring system." IEEE International Conference on Software Engineering & Service Science IEEE, 2015.

[13] Box, George E. P., et al. "Time Series Analysis: Forecasting and Control, 5th Edition." Journal of the Operational Research Society 22.2(2015):199-201.

[14] Oya, Kosuke, and H. Toda. "Dickey–Fuller, Lagrange Multiplier and Combined Tests for a Unit Root in Autoregressive Time Series." Journal of Time Series Analysis 19.3(2001):325-347.

[15] Konstantinides, Konstantinos, and K. Yao. "STATISTICAL ANALYSIS OF EFFECTIVE SINGULAR VALUES IN MATRIX RANK DETERMINATION." IEEE Transactions on Acoustics Speech & Signal Processing 36.5(2002):757-763.

[16] Clifford, Gari D, F. Azuaje, and P. Mcsharry. "Advanced Methods And Tools for ECG Data Analysis." Artech House, Inc. 2006.
[22] Oppenheim, Alan V., R. W. Schafer, and W. J. R. Buck. "Discrete-time signal processing." Electronics & Power 23.2(2009):157.