34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Sequential Markov Chain Monte Carlo Filter with Simultaneous Model Selection for Electrocardiogram Signal Modeling

Shwetha Edla, Narayan Kovvali, and Antonia Papandreou-Suppappola[1]

Abstract—**Constructing statistical models of electrocardiogram (ECG) signals, whose parameters can be used for automated disease classification, is of great importance in precluding manual annotation and providing prompt diagnosis of cardiac diseases. ECG signals consist of several segments with different morphologies (namely the P wave, QRS complex and the T wave) in a single heart beat, which can vary across individuals and diseases. Also, existing statistical ECG models exhibit a reliance upon obtaining *a priori* information from the ECG data by using preprocessing algorithms to initialize the filter parameters, or to define the user-specified model parameters. In this paper, we propose an ECG modeling technique using the sequential Markov chain Monte Carlo (SMCMC) filter that can perform simultaneous model selection, by adaptively choosing from different representations depending upon the nature of the data. Our results demonstrate the ability of the algorithm to track various types of ECG morphologies, including intermittently occurring ECG beats. In addition, we use the estimated model parameters as the feature set to classify between ECG signals with normal sinus rhythm and four different types of arrhythmia.**

I. Introduction

Electrocardiogram (ECG) signals are temporal recordings of electrical activity of the heart generated by the continuous depolarization and repolarization of cardiac cells. Detection of abnormalities in the ECG signals by manual examination can present a lot of difficulties, especially in long ECG recordings, such as Holter ECGs. However, construction of ECG signal models and use of the model parameters to perform automated disease classification can circumvent the need for manual annotation.

Several modeling methods using signal processing techniques have been proposed to describe ECG signals by means of orthonormal basis functions [1], autoregressive modeling [2] and linear prediction [3], among others. Delineation of the ECG signals was carried out to determine the position of the various ECG fiducial points and these were modeled using polynomial functions in [4], [5] and Hermite polynomials in [6], [7]. These approaches were limited by the fact that the shapes of the fiducial points depend upon individuals and their cardiac health, and thus a single mathematical representation cannot be used for their characterization. The dynamical nature of the ECG signals was exploited in [8], wherein the ECG signal was represented as a sum of five Gaussian functions. The parameters of the Gaussian functions were estimated in [9], [10] by constructing a statistical framework and using an extended Kalman filter (EKF) for parameter estimation. However, the model was not robust to initialization errors and non-linear solvers had to be used to determine the initialization parameters. Also, the presence of abrupt morphological changes in the ECG signal led to modeling errors. The representation of the ECG signal as a summation of multiple harmonics formed the basis for the dynamical model presented in [11], [12], wherein the model parameters represented by the harmonic coefficients and other cardiac signal attributes were estimated using both an EKF and a marginalized particle filter (PF). But this model relied upon the availability of *a priori* information such as the number of harmonics, and user-defined parameters such as the parameter noise variances, the values of which can vary greatly across different cardiac diseases as well as patients.

In order to avoid the aforementioned issues associated with existing ECG signal models, we propose a novel ECG model based on the sequential Markov chain Monte Carlo (SMCMC) filter [13] which can also simultaneously choose between different available models to represent the data. We adaptively delineate the dynamically varying ECG signal into windows, which represent a set of ECG samples during which a static model parameter assumption can hold good, in compliance with the SMCMC filter assumption that the model parameters are not time-varying. Within each window, the ECG signal is modeled using three different models, namely linear, quadratic and cubic polynomials, and the static model parameters represented by the polynomial coefficients are estimated sequentially. At the end of each window (determined adaptively based on the model likelihoods), the ECG signal is reconstructed using the estimates from each model weighted by the model probabilities. Using real ECG data, we show that the algorithm can successfully track (model) different types of ECG signals without requiring preprocessing steps or *a priori* information. We demonstrate the superior performance of our algorithm compared to the Gaussian ECG model [9], [10] for abruptly changing ECG morphologies, and finally use the estimated ECG signal parameters to distinguish between five different types of ECG signals.

II. Sequential Monte Carlo Markov Chain Filter and Model Selection

Particle filtering [14] is a sequential Bayesian approach, in which, the unknown state of a nonlinear dynamical system is determined by estimating its posterior probability density function (pdf), represented by a set of particles with associated weights, using the observed measurements or data. However, if the system or model parameters are assumed to be static,

This work was supported by the National Science Foundation (NSF) under Research Grant No. 0830799.

[1]Shwetha Edla, Narayan Kovvali and Antonia Papandreou-Suppappola are with the School of Electrical, Computer and Energy Engineering at Arizona State University in Tempe, AZ.

Emails: sedla@asu.edu, narayan.kovvali@asu.edu, papandreou@asu.edu

the sequential importance sampling (SIS) particle filtering approach has to be combined with MCMC methods (forming the SMCMC filter) in order to preserve the stability of the particle algorithm [13]. This is accomplished by rejuvenation of the particles, for every few samples of data that constitute a batch of size n, depending on the outcome of a rejuvenation test.

Assuming that at time step k, the posterior pdf $p(\mathbf{x}|\mathbf{Z}_k)$ of the unknown state $\mathbf{x}$ (not dependent on k because of the static assumption) is represented by the particle-weight pairs $\{(\mathbf{x}_j, w_{k,j})\}_{j=1}^{N_s}$, where $\mathbf{Z}_k = [\mathbf{z}_1, \ldots, \mathbf{z}_k]$ is the set of measurements upto time k and N_s is the number of particles, the first step in the SMCMC filter is to update the particle weights using SIS, as [13] (shown here for time step $k+n$),

$$w_{k+n,j} \propto p\left(\mathbf{z}_{k+1}, \ldots, \mathbf{z}_{k+n}|\mathbf{x}_j, \mathbf{Z}_k\right) w_{k,j}\,. \tag{1}$$

In order to prevent degeneracy and preserve stability, a rejuvenation test is performed next using the Kullback-Leibler (KL) distance [13]. If the KL distance falls below a threshold τ_1, in the final step, the independent Metropolis-Hastings (IMH) is used as the MCMC method of choice to generate a new set of particles and weights using the Gaussian proposal density, $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})$. The mean $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{x}}$ of the Gaussian proposal distribution are obtained using the particles values and their corresponding weights [13]. The rejuvenated particles are then used to estimate the unknown posterior pdf of $\mathbf{x}$.

When model selection is used along with the SMCMC filter, assuming that there are M models, $\{H_1, \ldots, H_M\}$, the posterior density is now given by [13],

$$p(\mathbf{x}|\mathbf{Z}_k) = \sum_{i=1}^{M} P(H_i|\mathbf{Z}_k) p(\mathbf{x}^i|\mathbf{Z}_k, H_i)\,, \tag{2}$$

where $\mathbf{x}^i$ is the state vector, $P(H_i|\mathbf{Z}_k)$ is the model probability, and $p(\mathbf{x}^i|\mathbf{Z}_k, H_i)$ is the posterior pdf for $\mathbf{x}$ given model i (for $i = 1, \ldots, M$). The model probability and posterior pdf are both updated sequentially using the model likelihood $p(\mathbf{z}_{k+1}|\mathbf{Z}_k, H_i)$ and SIS, respectively [13].

III. ECG Signal Modeling and Classification

In order to describe the dynamically varying ECG signal with changing morphologies, we use three different time-domain polynomial function models. The coefficients of these polynomials are the unknown parameters of the system, and are estimated using the SMCMC filter performing simultaneous model selection. We assume that the polynomial coefficients are static over short periods of time (windows). This assumption can be viable over segments of the ECG signal such as the P wave, QRS complex, ST segment, etc. We adaptively delineate the ECG data into windows (which is a set of samples over which model parameters are constant) using the model likelihood function. A window ends when the model likelihood falls below a certain threshold τ_2 and the models with their corresponding static parameters, no longer describe the data correctly.

A. State-space Model Framework

With the assumption that the signal parameters are constant within a given window l (where $l = 1, \ldots, L$, and L is the total number of windows), each sample of the ECG signal is represented as a polynomial of order M with $M+1$ unknown coefficients as,

$$z_{k,l} = \sum_{m=0}^{M} a_{l,m}\; t_{k,l}^{m} + v_k\,. \tag{3}$$

In the above equation, $z_{k,l}$ represents each ECG sample in window l and $t_{k,l} = k_l T_s$ is the discrete time at the kth sample in the lth window. The measurement noise v_k is assumed to be white Gaussian with zero mean and covariance R. The unknown coefficients of the Mth order polynomial form the state vector $\mathbf{x}_l$ in the lth window given by, $\mathbf{x}_l = [a_{l,0}\; a_{l,1}\; \ldots\; a_{l,M}]^T$. To differentiate between the state vector for each model, we assume that in general, for a polynomial of order M, only the first $M+1$ terms in $\mathbf{x}_l$ are non-zero. The state model is simply given by,

$$\mathbf{x}_{k,l} = \mathbf{x}_l\,. \tag{4}$$

In order to adaptively delineate the ECG data, the model likelihood $p(z_{k+1,l}|\mathbf{Z}_{k+1,l}, H_{i,l})$ is computed sequentially, compared to a threshold τ_2, and if it falls below this value, a new window $l+1$ is started by reinitializing the particles and their weights. At the end of the window, the signals are reconstructed and fit to the ECG data using the polynomial coefficient estimates at the last sample of the window since these denote the best estimates of the parameters.

B. Arrhythmia of the Heart and Classification

The electrical activity during each ECG beat originates from the depolarization of the pacemaker cells in the sinus node located at the top of the right atria, and propagates through all the chambers of the heart. Arrhythmia indicates a disturbance in the rate, regularity, site of origin, or conduction of the cardiac electrical activity [15]. In this work, we consider four types of arrhythmia for classification and comparison with normal sinus rhythm (N) signals, namely, the left bundle branch block (L) and right bundle branch block (R), which are two types of conduction block arrhythmias caused when the propagating electrical activity meets with unexpected delays along its path, and the ventricular escape (E) and junctional escape (j) beats, which are two types of escape rhythm arrhythmias caused when the electrical activity does not originate from the sinus node.

We utilize a simple Bayes maximum-likelihood (ML) classifier that uses the estimated model parameters as its features to perform classification. The features in our work are given by the reconstructed (estimated) signals that are computed using the parameter estimates for each ECG beat. In order to limit the dimension $N_{\boldsymbol{\beta}}$ of the feature vector $\boldsymbol{\beta}$, we form the feature vector using five points chosen from around the QRS complex and one additional point corresponding to the mean of the P wave samples, thus making $N_{\boldsymbol{\beta}} = 6$. The Bayes ML

classifier is based on maximizing the likelihood of the feature vector $p(\boldsymbol{\beta}|C_q)$ conditioned on the given class C_q (where $q = 1, \ldots, \mathcal{Q}$, and $\mathcal{Q}$ is the total number of classes), which is assumed to be a multivariate Gaussian distribution. Thus, the classifier output is given by, $C_q^* = \text{argmax}_q \ \log(p(\boldsymbol{\beta}|C_q))$. Here, the number of classes is $\mathcal{Q} = 5$.

IV. Simulation Results

In order to validate our algorithm and demonstrate its performance, we use real ECG data from the MIT-BIH arrhythmia database [16]. All signals are sampled at 360 Hz. We first process the ECG data using a lowpass filter to remove baseline wander and powerline interference as shown in [17]. Then the data is divided into beats using the peak location provided by the MIT-BIH arrhythmia database, since we wish to classify each data beat into the appropriate arrhythmia class. The parameters of each beat are estimated as described in Section III-A.

The tracking capability of our algorithm is shown in Fig. 1, that depicts the reconstructed ECG signal plotted against the original ECG data. Fig. 1(a), 1(c) and 1(d) show the results for normal sinus rhythm (N), ventricular escape (E) and junctional escape (j) type beats. It is observed that our algorithm can accurately reconstruct ECG signals of different morphologies, including those with abruptly occurring beats such as premature ventricular contraction (PVC), as seen in Fig. 1(b). Such type of signals were not tracked using the Gaussian ECG model presented in [9], [10].

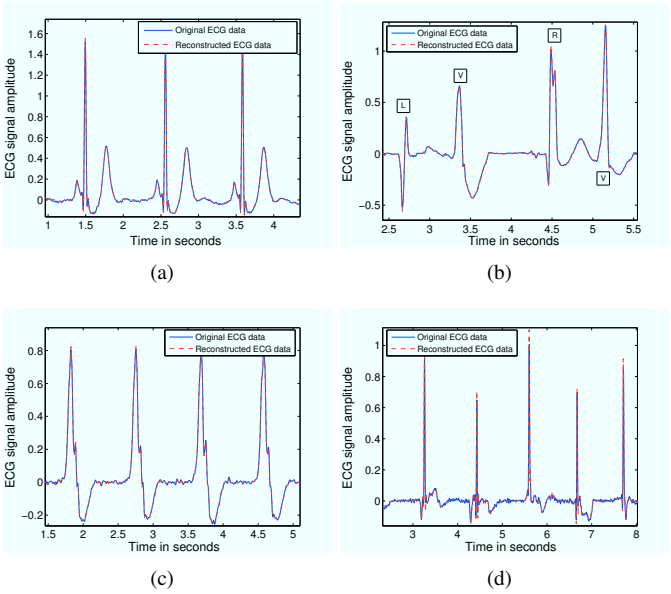


Fig. 1. Original and reconstructed ECG data using SMCMC filter and model selection for beat types with: (a) Normal sinus rhythm (N); (b) Left bundle branch block (L), Premature ventricular contraction (V) and right bundle branch block (R); (c) Ventricular escape (E); and (d) Junctional escape (j). The letters in the square boxes indicate the beat labels.

We demonstrate the superior performance of our algorithm by showing the tracking performance, and also comparing the estimation mean squared error (MSE) of the reconstructed signal obtained using our algorithm, to those obtained using the already existing nonlinear Bayesian framework for modeling ECG signals using Gaussian functions [9], [10]. The estimation MSE is calculated using the reconstructed signals obtained from both models and averaged over a number of Monte Carlo (MC) runs, as, MSE = $\frac{1}{N_r}\sum_{r=1}^{N_r}(\mathbf{Z}_{1:N} - \hat{\mathbf{Z}}_{1:N}^r)^2$, where N_r is the total number of MC runs ($N_r = 500$ is used for all MSE simulations in this work), $\mathbf{Z}_{1:N}$ represents the samples of the reference ECG signal, with N being the total number of data samples used, and $\hat{\mathbf{Z}}_{1:N}^r$ are the estimated signal samples obtained using the model parameter estimates for the rth MC run. The noise-free reference ECG signal is formed by averaging noisy real ECG beats of similar type obtained from the MIT-BIH arrhythmia database [16] and is shown in Fig 2(a). The better tracking performance of our algorithm can also be seen in Fig. 2(a) by comparing the plots showing the original and reconstructed signals using estimates from our SMCMC model and the Gaussian model of [9], [10] for a typical MC run. It can be observed that the Gaussian method does not track the data well and misses some of the fiducial points, such as the curves of the Q and T waves, among others, since it uses a phase-wrapping method to generate initial filter estimates. However, our algorithm can easily track the data by adaptively delineating it into windows and using the appropriate polynomial model. The superior performance of our algorithm in terms of estimation MSE can be seen in Fig. 2(b), which shows that our algorithm has lower estimation MSE when compared to the Gaussian ECG model.

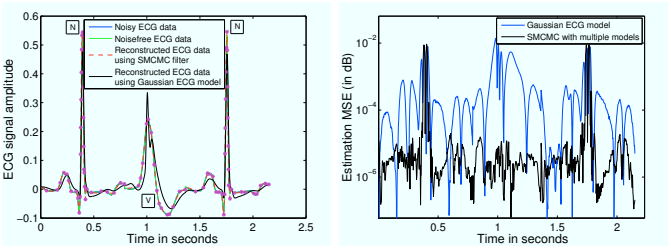
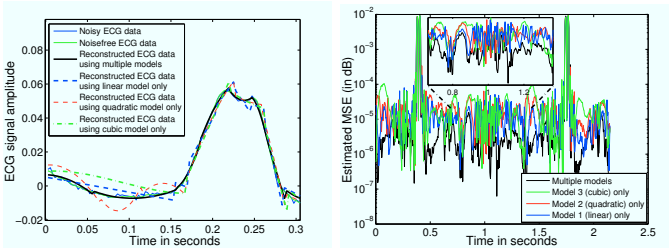


Fig. 2. (a) Original and reconstructed ECG data using SMCMC method with simultaneous model selection (The asterisks in the plot indicate the position in time at which a given window, during which model parameters are assumed to be constant, has ended) and Gaussian ECG model [9], [10]; (b) Comparison of estimation MSE using SMCMC method with simultaneous model selection and Gaussian ECG model [9], [10]. Note that the y-axis in the plot is logarithmic.

The leverage offered by multiple models is further substantiated in Fig. 3(a), wherein it is seen that the flexibility provided by the algorithm allows the best model to be selected to represent the data depending on the original signal, and provides better tracking results, compared to when a single model (either a linear, quadratic or cubic polynomial) is used without any model selection. For clarity, we only show the samples from the P wave of the first N type beat using the same ECG signal shown in Fig. 2(a). In addition, in Fig. 3(b), it is seen that the estimation MSE is lower while using multiple models when compared to the estimation MSE while using a single model.

Finally, we use the Bayesian ML classifier to classify between five different ECG signals as discussed in Section



(a) (b)

Fig. 3. (a) Original and reconstructed signal (shown only for the P wave portion) using multiple models with model selection and individual models with no model selection; (b) Comparison of estimation MSE using multiple models with model selection and individual models with no model selection. Inset shows a magnified version of plot between 0.7-1.3 seconds. Note that the y-axis in the plot is logarithmic.

III-B. The classification results are shown in the form of a confusion matrix in Table I, wherein, the rows and columns in the table indicate the true and the estimated class of the signals, respectively. The (u, v)th entry of the matrix gives the percentage of signals in class C_u that are classified to class C_v. The correct classification rates given by the diagonal entries of the matrix show that our algorithm achieves a fairly high classification rate, with an average correct classification rate of 98.5%. We also compare our correct classification rates with those presented in [6], [7] in Table II. In both these works, each delineated QRS complex was fitted with a Hermite polynomial and the feature vector consisting of 17 parameters was obtained. As seen from Table II, our classification results compare favorably with those results even with a fairly small feature set. In fact, our results for the correct classification rates of the j type beats are considerably better since our features include information about the P wave, which is noticeably absent in these beats [15].

TABLE I
CLASSIFICATION CONFUSION MATRIX FOR FIVE ECG SIGNAL TYPES

Class	N	L	R	E	j
N	97.9%	0	0	0	2.1%
L	0	99.5%	0	0.5%	0
R	0	0	97.9%	2.1%	0
E	1.8%	0	0	98.2%	0
j	1.2%	0	0	0	98.8%

TABLE II
COMPARISON OF CLASSIFICATION RESULTS

Class Type	Classification Rate [6]	Classification Rate [7]	Classification Rate (New)
N	98.1%	97.8%	97.9%
L	97.0%	96.6%	99.5%
R	94.0%	99.0%	97.9%
E	90.0%	96.0%	98.2%
j	–	90.5%	98.8%

V. CONCLUSIONS

In this paper, we presented a novel method for ECG modeling and parameter estimation using the SMCMC filter with simultaneous model selection. We showed that our algorithm can track different ECG morphologies and beat types without requiring *a priori* information about the data. Our results also demonstrated a superior estimation MSE performance when compared to the Gaussian ECG model [9], [10] for tracking ECG signals with abrupt changes in morphologies. In addition, an average correction classification rate of 98.5% was obtained for classifying five different types of ECG signals using the model parameters.

REFERENCES

[1] L. Sornmo, P. O. Borjesson, M.-E. Nygards, and O. Pahlm, "A method for evaluation of QRS shape features using a mathematical model for the ECG," *IEEE Transactions on Biomedical Engineering*, vol. 28, pp. 713–717, Oct. 1981.

[2] J. P. Marques de Sá and C. Abreu-Lima, "Optimal autoregressive modeling of ECG signals," in *Computers in Cardiology*, 1987, pp. 147–149.

[3] L. Kang-Ping and W.H. Chang, "QRS feature extraction using linear prediction," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 10, pp. 1050–1055, Oct. 1989.

[4] R. Borsali, A. Nait-Ali, and J. Lemoine, "ECG compression using ensemble polynomial modeling: Comparison with the wavelet-based technique," *Biomedical Engineering*, vol. 39, pp. 138–142, May 2005.

[5] S. Jokic, V. Delic, Z. Peric, S. Krco, and D. Sakac, "Efficient ECG modeling using polynomial functions," *Electronics And Electrical Engineering*, vol. 110, no. 4, 2011.

[6] S. Osowski and T. H. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 1265–1271, Nov. 2001.

[7] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 582–589, Apr. 2004.

[8] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, pp. 289–294, Mar. 2003.

[9] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A nonlinear bayesian filtering framework for ECG denoising," *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 2172–2185, Dec. 2007.

[10] O. Sayadi and M.B. Shamsollahi, "ECG denoising and compression using a modified extended Kalman filter structure," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 9, pp. 2240–2248, 2008.

[11] J. McNames and M. Aboy, "Statistical modeling of cardiovascular signals and parameter estimation based on the extended Kalman filter," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 119–129, Jan. 2008.

[12] S. Kim, L. Holmstrom, and J. McNames, "Multiharmonic tracking using marginalized particle filters," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2008, pp. 29–33.

[13] D. S. Lee and N. K. K. Chia, "A particle algorithm for sequential Bayesian parameter estimation and model selection," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 326–336, 2002.

[14] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[15] M.S. Thaler, *The Only* EKG *Book You'll Ever Need*, Lippincott Williams & Wilkins, 2010.

[16] 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, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation Electronic Pages*, vol. 101, pp. 215–220, Jun. 2000.

[17] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1196–1206, Jul. 2004.