

Cardiac Arrhythmia Detection by Parameters Sharing and MMIE Training of Hidden Markov Models

Carlos S. Lima, Manuel J. Cardoso

Department of Industrial Electronics of University of Minho, Campus de Azur em, Guimar es, Portugal
carlos.lima@dei.uminho.pt

Abstract—This paper is concerned to the cardiac arrhythmia classification by using Hidden Markov Models and Maximum Mutual Information Estimation (MMIE) theory. The types of beat being selected are normal (N), premature ventricular contraction (V), and the most common class of supra-ventricular arrhythmia (S), named atrial fibrillation (AF). The approach followed in this paper is based on the supposition that atrial fibrillation and normal beats are morphologically similar except that the former does not exhibit the P wave. In fact there are more differences as the irregularity of the RR interval, but ventricular conduction in AF is normal in morphology. Regarding to the Hidden Markov Models (HMM) modelling this can mean that these two classes can be modelled by HMM's of similar topology and sharing some parameters excepting the part of the HMM structure that models the P wave. This paper shows, under that underlying assumption, how this information can be compacted in only one HMM, increasing the classification accuracy by using MMIE training, and saving computational resources at run-time decoding. The algorithm performance was tested by using the MIT-BIH database. Better performance was obtained comparatively to the case where Maximum Likelihood Estimation training is used alone.

1. INTRODUCTION

Electrical instability of the heart, which can be identifiable in the ECG, leads to an abnormal synchronized contraction sequence reducing pumping efficiency. This phenomenon named arrhythmia can be classified as frequent or infrequent (sporadic). Infrequent arrhythmias can be evaluated by long-term ambulatory ECG monitoring (Holter), which produces a quantity of beats greater than 10^5 . The physician has now the very hard task of visual examination of this huge amount of data to find possible abnormal beats that in some cases are morphologically similar (but not equal) to normal beats. In spite of time consuming and boring work the task of repetitively analyzing data that in the most cases are normal beats increases the human error by subjective reasons. This subjective error increases significantly when ectopic pulses are very similar with the most frequent normal pulses, which occurs for atrial fibrillation. Usually ventricular arrhythmias are more visible

given the QRS complex is the most dominant part of the ECG. Hence, an automatic method for beat classification of a Holter register eventually separating some beats hard to classify, for posterior visual inspection, would be very useful.

Hidden Markov models have been successfully applied to pattern recognition problems in applications spanning automatic speech recognition [1], image segmentation [2], ECG modeling [3] and cardiac arrhythmia analysis [4].

The approach followed in this paper tries to take advantage of the similarities between normal and atrial fibrillation beats to improve the classifier performance by using MMIE training, in a single model/double class framework, which is similar of having two different models sharing the most parameters. This approach saves computational resources at run-time decoding and improves the classification accuracy of very similar classes by using MMIE training. The idea is that if two classes have some state sequence similarities and the main morphological differences occur only in a short time slice, then setting appropriately internal state model transitions can model the differences between classes. These differences can be more efficiently emphasized by taking advantage of the well known property of MMIE training of HMM's, which typically makes more effective use of a small number of available parameters [7]. By this reasoning the selected decoding class can be chosen on the basis of the most likely state sequence, which characterizes the most likely class. Classes morphologically not too similar are modeled by different HMM's by using MLE training alone since one property of MMIE training is that training data for which the probability of being generated by one existing model is much greater than the probability of being generated by anyone of the others, have negligible contribution to the reestimated values.

II. ECG FEATURES EXTRACTION

ECG observations were obtained from the segmentation of the original signal with straight line segments which goal is to decrease the amount of linear redundancy, as described in [3]. In [3] it is suggested for features a bi-dimensional vector where the components are respectively the amplitude of the starting point and the duration of the line segment. However, as reported in [5], these features are very sensitive to baseline wander, DC drift and heart rate variation. DC drift can be cancelled by using differential amplitude between the starting and ending points, and heart rate variability can be attenuated

by normalizing the line segment duration by the R-R interval, as reported in [5]. Therefore we adopted the features suggested in [5]. The R-R interval is computed by using the well known Gritzali algorithm [6], which is also used jointly with a valley detector for beat synchronization. As the used HMM's are connected in a left to right order, synchronization of the cardiac cycle according to the initial state probability is required especially for training purposes. For decoding this synchronization is only necessary for the first cardiac cycle since HMM's are provided with a feedback transition from the last to the first state.

III. HIDDEN MARKOV MODELS

A. Model Structure

In the pattern recognition paradigm each class of beat is represented by a separate model and after decoding, the class for the which the probability (likelihood) of occurrence is greater is selected. Since the ECG is characterized by time sequence waves occurring almost always in the same order which reflects the sequential activity of the cardiac conduction system an HMM structure where the states are connected in a left-to-right order was adopted. In [3] it is shown that a full connected HMM is eventually more appropriate for HMM modeling since the beat sequence reproduced by this kind of HMM is almost perfect. However, it is well known that classification in the pattern recognition paradigm does not need necessarily of modeling all the class features, so though a left-to-right model may not be the more adequate, it is structurally appropriated from an heuristic point of view and can capture the most relevant features concerned to classification purposes. Figure 1 shows the model structure for the atrial fibrillation and normal beats, where $a_{i,j}$ stands for transition probability from state i to state j . Our reasoning is based on the assumption that an AF beat is similar to a normal beat without the P wave which can be modeled by a transition probability that not pass through the state which models the P wave. At the end of the decoding stage the recognized class can be selected by searching (backtracking) the most likely state sequence. This structure can be seen as two separate HMM's sharing the most parameters. This parameter sharing procedure is justified by the fact that ventricular conduction is normal in morphology for AF beats, and we intend to use a limited amount of parameters, just the pdf associated with the transitions from state 5 to states 6 and 7, state 6 to itself and to state 7 to reinforce the discriminative power among classes. The separation between these two classes can be increased by using an efficient discriminative training as MMIE obtained on the basis of the parameters associated with the intra-class differences, just those above mentioned. It is very important to note that this approach reinforces the HMM distance among different model structures while the distance of HMM's in the same structure (those that share parameters) are obviously decreased. However, it is believed that an appropriate discriminative training can efficiently separate the classes modeled by the same HMM. Although a recognition system fully trained by using the MMIE approach can be more effective it surely needs a much degree of computational requirements in both training and run time decoding.

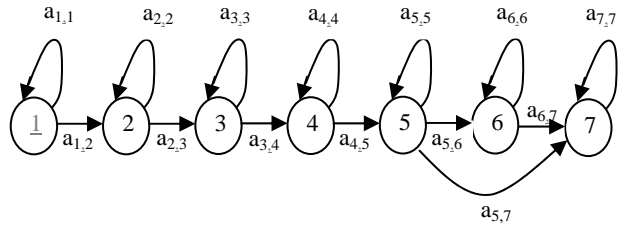


Figure 1. HMM topology adopted for modelling normal (N) and atrial fibrillation (AF) beats.

States from 1 to 7 are concerned to the ECG events R, S, S-T, T, T-P, P, P-R.

The model structure adopted for PVC beats is similar to the shown in figure 1 but without state 6, since the P-wave is not present. AF and PVC beats have the similarity of do not exhibit the P wave, however they are morphologically very different, therefore it is not plausible that they can share a significant amount of model parameters.

B. Probabilistic model of observations

The output probability density function, which defines the conditional likelihood of observing a set of features when a transition trough the model takes place, is usually a multivariate Gaussian mixture, so the probabilistic model assigned to observation vectors is a bi-variate Gaussian probability density function since the observation vectors have only 2 components. The components of observation vectors are assumed to be independents and identically distributed (iid) hence the joint likelihood occurrence is given by the product of two Gaussian functions. These probability density functions are associated with transitions which configures a Continuous Density Hidden Markov Model (CDHMM) Mealy machine and are given by

$$f(y/u_t) = \sum_{i=1}^C b_{u_t,i} G(y_t, \mu_{u_t,i}, \Sigma_{u_t,i}) \quad (1)$$

where $G(\dots)$ stands for bi-variate normal distribution with mean vector and covariance matrix for the i^{th} mixture component and transition u_t given respectively by $\mu_{u_t,i}$ and $\Sigma_{u_t,i}$. As the components of observation vector are assumed iid $G(\dots)$ function in equation (1) is simply the product of two Gaussian functions. The mixture coefficients $b_{u_t,i}$ satisfy, for each transition u_t , to

$$\sum_{i=1}^C b_{u_t,i} = 1 \quad (2)$$

so that, equation (1) is a probability density function.

In our experiments the observations were modelled by three components in the Gaussian mixture ($C=3$) in order to best fit the data with multimodal distributions.

C. Training Procedure

The Estimation of HMM parameters from a set of representative training data can be done by using the Baum-Welch algorithm which is based on the decoding of all the possible state sequence, or alternatively by using the Viterbi algorithm which is based on the most likely state sequence [1]. Since the HMM structure shown in figure 1 can model 2 different classes on the basis of the most likely state sequence, the Viterbi algorithm seems to be more appropriate for this kind of decoding strategy, once that after decoding, the most likely state sequence can be known by an appropriate backtracking procedure.

The frame state allocation concerned to the ECG events described in the first paragraph after figure 1 can be forced by setting (to one) the initial probability of the first state in the initial state probability vector and resetting all the other initial state probabilities, and also synchronizing the ECG feature extraction to begin in the R wave. This kind of synchronization is needed for this HMM topology where the initial state must be synchronized with the R wave, otherwise the assumption that state 6 models P-wave can not be true. We observed this evidence in our experiments. However if a back transition from the last to the initial state is added this synchronization is necessary only for the first ECG pulse decoding. The synchronization between ECG beats and the HMM model is facilitated by the intrinsic difference between the last and first state, since the last state models an isoelectric segment (weak signal) while the first state models the R wave which is a much strong signal. In other words if the HMM is in state 7 modeling an isoelectric segment the happening of a strong R wave tends to force a transition to state one which helps in model/beat synchronization.

The model for premature ventricular contraction beats is trained by using the conventional MLE procedure in the Viterbi framework, which goal is to maximize iteratively the following probability density function

$$f(Y/\lambda) = f(Y/S, \lambda)P(S/\lambda) \quad (3)$$

where Y is the observation sequence, S the most likely state sequence and λ the set of HMM parameters. The model reestimation formulas can be found in [1]. This usual parameter estimation technique attempts to make the models fit the training data as well as possible.

Another reasonable training objective would be to maximize the mutual information between the training sequence and the corresponding observation sequence given the set of existing models. This training criterion leads to the maximization of the following probability density function

$$f(\lambda/Y) = \frac{f(Y/S, \lambda)P(S/\lambda)}{\sum_{\lambda'} f(Y/S, \lambda')P(S/\lambda')} \quad (4)$$

The most important thing that can be immediately observed from this objective function is that maximizing it is equivalent to enforcing discrimination against all competing models. The main difficulty associated with the use of MMIE estimation in HMM's is the non-existence of closed-form reestimation

formulas similar to those available for MLE. So, a common solution is to resort to some form of gradient descent or alternatively relying on efficient reestimation techniques which main virtue is not as much their proofs of guaranteed convergence as their effectiveness in practice given that convergence is reached in a few (typically less than 10 and often 2 or 3) iterations. One such technique was proposed in [8] for discrete distributions and adapted in [7] for continuous distributions and was selected to be used in the ambit of this paper.

As different state sequences model different classes in the same HMM a suited training procedure can be used, taking into consideration that this model structure is similar to a structure with two HMM's sharing a significant amount of parameters. The approach followed in this paper was to compact this representation in only one HMM. The adopted training strategy must accommodate both the MMIE training and parameter sharing, or in other words an MMIE training procedure in only one HMM platform with capabilities to model two classes must be required. This compromise was obtained by estimating the shared parameters in the MLE sense. This procedure emphasizes that the shared parameters can be estimated on the basis in which the data fits best the model. For this propose a set of 20 normal beats was presented to the HMM structure shown in figure 1 with $a_{5,7}$ set to zero which means that the pdf parameters associated with this transition were not trained. The Viterbi algorithm was used for training and testing purposes [1]. Then 10 AF beats and 10 N beats were presented for training, in the MMIE sense, the parameters not shared by the two classes, just the ones associated with the transitions $a_{5,7}$, $a_{5,6}$, $a_{6,6}$, $a_{6,7}$. All the other parameters are shared between the two classes and are not updated at this phase. Associated to each transition are 15 coefficients, three mixture coefficients; three mean vectors and three diagonal covariance matrices for two iid vector components. In this way this HMM can model efficiently "on average" beats morphologically similar to normal beats and additionally was specialized in distinguishing normal from AF beats. Our results seem to confirm this reasoning.

Probability state transitions $a_{5,7}$ and $a_{5,6}$ are concerned to the *a priori* beat probability since they serves as a switch between both classes modeled by this HMM. Hence for a non-biased model they must be numerically equal, which means that given an unknown beat the *a priori* probability of being an N beat is the same that of being an AF beat. Therefore these two model parameters must be set as

$$a_{5,6} = a_{5,7} = \frac{1 - a_{5,5}}{2} \quad (5)$$

in order to set the transition probability from state 5 unitary, as required for all states, and where $a_{5,5}$ was trained for normal beats and was not updated for AF beats since it is a shared parameter.

The PVC model was trained with 20 premature ventricular beats in the standard way, by using the Viterbi algorithm [1].

Good initial parameter estimates are very important in reaching the globally optimum parameter estimates. This was accomplished by manual segmentation of two examples of each beat type. The output mean values were initialised as the

sample means of the associated segments computed for each mixture component by the K-means algorithm.

IV. EXPERIMENTAL RESULTS

Experimental results were evaluated by using the MIT-BIH database. In order to show the effectiveness of the proposed algorithm we compared the performance of the algorithm relatively to the most common case where each class has its own HMM. Normal beats were modelled by the HMM topology shown in figure 1 where the transition from state 5 to state 7 was removed. The HMM training procedure used in this framework was the MLE with the Viterbi algorithm.

The testing set contains the 106, 119 and 123 records of the MIT-BIH arrhythmia database and the 04043 record of the MIT-BIH atrial fibrillation database. The training data of N, V and AF beats was taken respectively from the 100, 116 and 04126 records, which means that data for training and testing purposes was obtained from different patients, which is normally known as patient-independent analysis. An experimented cardiologist selected 10 good examples of AF cardiac cycles from the first two AF episodes of the 04126 record. AF testing data was selected by the same cardiologist, as good examples from the 1st and 12nd AF episodes of the 04043 record where 516 AF cardiac cycles were selected for this purpose. The signals were previously denoised using wavelet based filter and the baseline signal removal has been eliminated. Additionally corrective training was performed for both MMIE and MLE training. Tables 1 and 2 show the results in a confusion matrix form for the cases of MMIE and MLE alone training.

Table 1 – The confusion matrix associated to MMIE training

	V	N	AF	FP	Total	Pr+
V	963	1	3	23	990	0.98
N	2	4563	0	17	4582	1
AF	4	0	512	51	567	0.91
NR	5	1	7			
Total	974	4565	522	91	6139	
Sensitivity	0.99	1	0.98			

Table 2 – The confusion matrix associated to MLE training

	V	N	AF	FP	Total	Pr+
V	964	3	0	34	1001	0.97
N	1	4531	33	12	4577	1
AF	7	29	480	67	583	0.88
NR	4	13	11			
Total	976	4576	524	113	6161	
Sensitivity	0.99	0.99	0.92			

The results confirm that confusion among classes of beats morphologically similar can be strongly reduced by using

MMIE training, as shown by the confusability decreasing between N and AF beats from table 2 to table 1. Additionally no other significant differences appear, at least in this limited experimental set.

V. DISCUSSION AND FUTURE WORK

This paper suggests that robustness concerned automatic cardiac diagnosis can be increased by using MMIE training of HMM's, which model beat types of similar morphology. The idea is that it can be more effective specialising the HMM's in learning the differences between beats of similar morphology than learning the probability distributions that fit best the training data. Although the experimental results need to be extended specially in the number of classes to be recognized, which certainly increases the confusability among beat classes, they support the approach, as shown by the confusability decreasing between N and AF beats from table 2 to table 1. This approach can also be extended to the case of Atrial Flutter (AFL), where the P-wave repetitions along the cardiac cycle can also be viewed as a preferred (most likely) state sequence along the HMM, while the QRS is normal.

VI. REFERENCES

- [1] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE vol. 77, pg. 257-286.
- [2] Choi, H. and Baraniuk, R. G. (2001). Multiscale image segmentation using wavelet-domain hidden Markov models. IEEE Trans. Image Process., vol. 10, no. 9, pp 1309-1321.
- [3] Koski, A., (1996). Modelling ECG signals with hidden Markov models. Artificial Intelligence in Medicine, Vol. 8, pp. 453-471.
- [4] Coast, D. A., Stern, R. N., Cano, G. G., and Briller, S. A. (1990). An approach to cardiac arrhythmia analysis using hidden Markov models. . IEEE Trans. on Biomedical Engineering, Vol. 37, No. 9, pp. 826-836.
- [5] Cheng, W. T. and Chan, K. L. (1998). Classification of Electrocardiogram using hidden Markov models. Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 20, No. 1, pp. 143-146.
- [6] Gritzali, F. (1988). Towards a generalised scheme for QRS detection in ECG waveforms. Signal Process. Vol. 15, pp. 183-192.
- [7] Normandin, Y. (1991). Hidden Markov Models, Maximum Mutual Information Estimation, and the Speech Recognition Problem. Ph. D. Thesis, McGill University, Montreal.
- [8] Gopalakrishnan, P. S., Kanevsky, D., Nádas, A. and Nahamoo, D. (1989). A Generalisation of the Baum Algorithm to Rational Objective Functions. Proc. ICASSP-89, paper S12.9, Glasgow.