Analysis of the Observation Sequence Duration of Hidden Markov Models for QRS Complex Detection in Single-Lead ECG Recordings

Nelson F. Monroy, Miguel Altuve

Faculty of Electrical and Electronic Engineering, Pontifical Bolivarian University, Bucaramanga, Colombia

Abstract

Hidden Markov models (HMM) are useful tools to characterize the dynamics of observation sequences. However, the duration of the observation sequence to characterize is often chosen empirically, based on a priori information. In this paper, we used a grid search approach to optimize two hyperparameters of an HMM-based QRS complex detector, namely, the duration of the observation sequence to be characterized and the adaptive decision threshold that compares the difference of log-likelihoods of observations from two competing HMM. To assure the reproducibility of the results, we have run the optimization process ten times using different random seeds per realization. Using the ECG signals from the MIT-BIH Arrhythmia database, we have found an optimal adaptive decision threshold of 60% but the optimal value of the observation sequence duration varies from 100 ms to 120 ms for different realizations of the optimization process. Consequently, the different parameters of the HMM-based QRS complex detector found in each realization do not always lead to the same detection performance. Using the optimal values, the detector achieves a sensitivity of 96.53% and a positive predictivity of 98.16%.

1. Introduction

Machine learning, a branch of artificial intelligence, has become a leading tool in prediction, classification and clustering tasks in a plethora of areas, ranging from marketing campaigns to virtual assistants. In particular, in medicine, machine learning has proven useful in diagnosing cancer [1] and in predicting epidemic outbreaks [2], and is gaining popularity to become the backbone of clinical decision support systems [3, 4]. The success of a machine learning technique in a specific application depends on many factors, including those related to the technique, such as the choice of a particular approach and the parameters and hyperparameters associated with it, and those related to the application, such as the amount and quality of the data. Concerning the factors associated with the machine learning approach, the parameters of the model are estimated in a training phase from the data using an optimization algorithm, while the hyperparameters, on the other hand, cannot be estimated from the data and are, thus, established before the training phase using heuristic rules and prior knowledge. However, optimization techniques such as grid search, random search, Bayesian optimization, and evolutionary algorithms have also been used to find out hyperparameters [5–7].

A Hidden Markov model (HMM) is a powerful and very simple machine learning approach used to model a Markovian stochastic process by learning the dynamics of observation sequences (examples) generated by the process itself. An HMM is characterized by M states and the set of parameters $\lambda \triangleq \{a_{ij}, b_j, \pi_i\}$, where a_{ij} is the transition probability between states i and j ($a_{ii} \neq 0$), b_j is the probability of emission of observations of state j, π_i is the probability of state i being the initial state. The observation sequence of duration T is represented by $O = \{o_1, o_2, \ldots, o_T\}$. HMM have been applied in classification and clustering tasks in different domains given their flexibility and robustness for the treatment of univariate and multivariate, discrete or continuous, observations.

HMM are commonly used to characterize the dynamic of observation sequences by two distinct ways. The first, by analyzing the optimal state sequence associated with the observation sequence (Problem #2 in [8]), such as in speech recognition [9] and DNA sequence analysis [10] problems, and the second, by analyzing the probability of the observation sequence given the model (Problem #1 in [8]), such as in apnea-bradycardia detection [11] and ischemic episodes classification [12]. In the last case, besides the problems properly associated with the model (learning λ and choosing M), the selection of features (or feature vector) whose dynamics must be modeled, the amount of data needed for training and the duration of the observation sequence remain as problems generally addressed empirically. For instance, in the detection of apnea-bradycardia episodes in preterm infants, the duration of the observation sequence was related to the time measured from the beginning of the bradycardia to the peak in the RR interval time series [11], but a better detection performance could be obtained with

another choice, not necessarily related to a physiological phenomenon.

When selecting the observation sequence duration, one expects that the observation sequence must contain the dynamic to be modeled, but this selection can be particularly difficult and can affect the classification/detection performance. For instance, since the length of the observation sequence is related to the number of frequency components one can represent, a short duration can enhance the characterization of high-frequency components of the data but reduce low-frequency ones, and vice versa. In addition, the model can be biased by using a priori information (mostly related to the physical phenomenon) that most of the time is not considered by the model or is meaningless. In this paper, we are particularly interested whether the observation sequence duration of an HMM can affect the detection performance of QRS complexes in single-lead ECG recordings. Using the grid search approach, two hyperparameters of the QRS complex detector based on HMM were optimized: i) the duration of the observation sequence and ii) the adaptive detection threshold. The optimization looks to minimize the distance to perfect detection in an ROC curve.

2. Methods

2.1. HMM-based QRS complex detector

ECG signals were preprocessed to reduce noise and artifacts and enhance the temporal evolution (dynamics) associated with QRS complexes. To this end, ECG signals were band-pass filtered with sixth-order high-pass and low-pass Butterworth filters with cut-off frequencies 5 Hz and 30 Hz, respectively. Baseline wandering and P- and T-waves contributions were reduced in the ECG signal with these filters.

Like [11], two HMM were used to represent two different observation sequence dynamics in the preprocessed ECG signal: i) λ_{QRS} to model the *presence* of a QRS complex in an observation sequence O_{QRS}, and ii) λ_{NQRS} to model the *absence* of a QRS complex in an observation sequence O_{NQRS}. We used continuous density HMM with Gaussian distribution as the probability of emission of observations for a given state, i.e. b_j is represented as a Gaussian distribution, $b_j(\mu_j, \sigma_j^2)$, with mean μ_j and variance σ_j^2, where μ_j corresponds to the barycenter of state j in the observation space. Each Gaussian distribution represents a portion of the dynamic range of the observation and is tied to a specific state. Therefore, given the low dynamic range of the observation sequences O_{QRS} and O_{NQRS}, we used $M = 3$ states for each HMM. To increase the dynamic range and avoid having Gaussian distributions with narrow variance ($\sigma_j^2 \rightarrow 0$), we multiplied O_{QRS} and O_{NQRS} by 10 and 5, respectively. The values of these multiplication factors were chosen by visual inspection of the Gaussian distributions to

get well-defined bell curves for $b_j(\mu_j, \sigma_j^2)$. In addition, a_{ij} and π_i were initialized with uniform probabilities.

Following the recommendations of the ANSI/AAMI EC38:1998 standard [13], the HMM-based detector was trained in a learning period represented by the first five minutes of the records; the remainder of the records were used to test its detection performance. In this sense, two training datasets were created: one containing examples of O_{QRS} and the other containing examples of O_{NQRS}. Using the annotations of QRS complexes provided in the database, examples of O_{QRS} of length T, centered at the annotation marks, were randomly chosen from the first five minutes of the records. Examples of O_{NQRS} of length T were also randomly chosen from the first five minutes of the records but without having a QRS annotation mark within 150 ms. The test dataset is composed of 48 excerpts of the preprocessed ECG signals, taken from minute 5 to the end of the record.

During the learning phase, λ_{QRS} and λ_{NQRS} were estimated using the Baum-Welch algorithm [14] on the appropriated training dataset. Training is achieved when the log-likelihood $\mathcal{L} = \log P(O|\lambda)$ converges to a maximum value or when $\frac{\mathcal{L}^{it} - \mathcal{L}^{it-1}}{\mathcal{L}^{it}} < 0.01$, where it is the current iteration of the Baum-Welch algorithm. During the test phase, the detection of a QRS complex at time t is signaled when the difference of log-likelihoods of producing a new observation sequence O by each model (λ_{QRS} and λ_{NQRS}) is greater than an adaptive threshold:

$$z(t) = \frac{\mathcal{L}(t)_{QRS} - \mathcal{L}(t)_{NQRS}}{T} > \alpha w(t), \qquad (1)$$

where $w(t)$ is the average value of z(t) during the last 15 s and α is a parameter that must be optimized. A sliding window of length T was used to extract the observation sequences to be passed through the models. The shift of this window was set to 30 ms. The HMM-based QRS detector was implemented in MATLAB using the HMM toolbox by K. Murphy [15].

2.2. QRS complex detection performance

The detection approach was validated using the first ECG channel of the MIT-BIH Arrhythmia database [16,17]. This database contains 48 half-hour excerpts of ECG recordings, digitized at 360 samples per second with 11-bit resolution over a 10-mV range, with reference annotations of QRS complex positions.

Sensitivity (Se) and positive predictivity ($+P$) were used as performance metrics of the QRS complex detector. These metrics were defined as $Se = \frac{TP}{TP+FN}$ and $+P = \frac{TP}{TP+FP}$, where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. A signaled detection is considered a TP if it lies within

Table 1. Best detection performances, measured as the lowest DDP, on each realization of the optimization process.

Realization	T (ms)	α	Se	$+P$	DDP
1	100	0.6	0.95	0.97	0.047
2	100	0.6	0.96	0.98	0.039
3	110	0.6	0.96	0.98	0.040
4	110	0.6	0.96	0.98	0.040
5	100	0.6	0.96	0.97	0.043
6	110	0.6	0.95	0.98	0.043
7	120	0.6	0.95	0.98	0.045
8	110	0.6	0.96	0.97	0.040
9	110	0.6	0.95	0.98	0.047
10	120	0.6	0.96	0.96	0.047

150 ms from a reference annotation, otherwise, it is considered an FP.

2.3. Optimization of hyperparameters

The HMM-based detector is composed of several parameters and hyperparameters. For each model, the Baum-Welch algorithm was used to find the set of parameters $\lambda \triangleq \{a_{ij}, b_j, \pi_i\}$ in the learning period, using the training datasets. However, M, the multiplication factor to avoid narrow Gaussian distributions for b_j, and the number of samples in $z(t)$ to compute $w(t)$ were chosen experimentally in order to provide the best detection performance.

The length of the observation sequence T and the multiplication factor α can be considered hyperparameters of the detection approach. A grid search method was used to find their optimal values by minimizing an objective function defined as the Euclidean distance to perfect detection (DPD) in an ROC curve: $DPD = \sqrt{(1 - Se)^2 + (1 - +P)^2}$. The search spaces were set to [50, 200] ms for T and [0, 1] for α.

The optimization process was repeated 10 times to ensure the reproducibility of the results. In this sense, the random seed was initialized to give different results in each realization of the optimization process.

3. Results

Table 1 shows the detection performances obtained at the lowest DDP, for each realization of the optimization process. We can observe that the optimal value of α is 0.6, but an optimal value of T ranging from 100 to 120 ms is obtained on each realization. This means that an observation sequence of duration 100 to 120 ms, centered at the R-wave peak in the band-pass filtered ECG signal, contains the greatest amount of information employed by both HMM to distinguish the two different dynamics (presence and absence of a QRS complex).
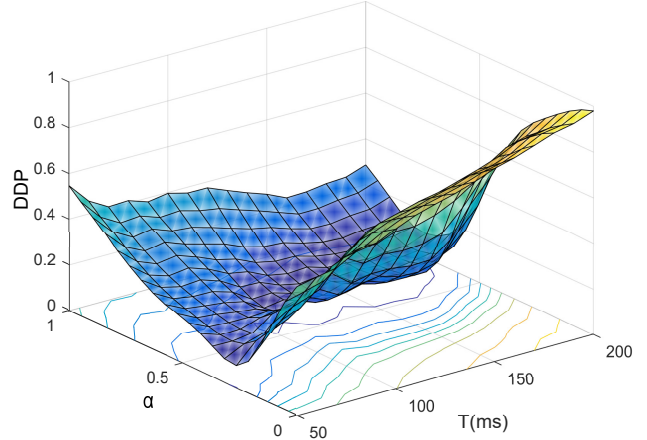

Figure 1. DDP averaged over 10 realizations of the optimization process.

Table 2. Average of the detection performances over the 10 realizations of the optimization process for $\alpha = 0.6$.

T (ms)	Se	$+P$	DDP
50	0.97	0.78	0.220
60	0.97	0.82	0.178
70	0.97	0.83	0.167
80	0.96	0.84	0.157
90	0.96	0.86	0.138
100	0.95	0.92	0.087
110	0.93	0.92	0.100
120	0.94	0.95	0.074
130	0.93	0.96	0.074
140	0.91	0.94	0.100
150	0.92	0.97	0.078
160	0.92	0.96	0.082
170	0.92	0.96	0.087
180	0.91	0.96	0.091
190	0.89	0.95	0.116
200	0.78	0.94	0.221

Figure 1 shows the DDP, averaged over the 10 realizations of the optimization process, for each value of T and α. Similarly, Table 2 shows the detection performances, averaged over the 10 realizations of the optimization process, for the optimal value of α (i.e. $\alpha = 0.6$) for each T. We can see that the average DDP curve is composed of many local minimums, with a global minimum at $T = 120$ ms and $\alpha = 0.6$.

4. Discussion

As shown in Figure 1, the detection performance varies for different combinations of T and α. In addition, we have observed that different detection performances were obtained for different realizations of the optimization pro-

cess (see Table 1). Indeed, the best detection performances were achieved for the same value of α (i.e. $\alpha = 0.6$) but for different durations of the observation sequence ($100 < T < 120$ ms). This can be a consequence of the susceptibility of the HMM to characterize the dynamics of the presence and absence of a QRS complex from the preprocessed ECG signal since each iteration of the optimization process produces a different set of HMM parameters (λ_{QRS} and λ_{NQRS}). The optimal value of T range from 100 to 120 ms, which is within the limits of the duration of QRS complexes from a physiological point of view (70 to 110 ms [18]).

5.   Conclusion

Hyperparameters affect the performance of machine learning approaches. In this work, we have used a grid search method to find the optimal values of two hyperparameters of an HMM-based QRS complex detector, namely, the duration of the observation sequence and the percentage of an adaptive threshold to compare the difference in log-likelihoods. The HMM-based detector was trained to differentiate two dynamics in the band-pass filtered ECG signal, i.e. the presence and absence of a QRS complex. By repeating the optimization process ten times, with different random seed initializations per realization, we have found that the set of parameters of HMM produces not only different detection performances but also different optimal values of the duration of the observation sequence (from 100 to 120 ms). In addition, we have found that an adaptive threshold that considers the 60% of the average of the difference in log-likelihoods for 15 s is optimal to maximize the detection performance. Using the optimal values, the best QRS complex detection performance was 96.53% of sensitivity and 98.16% of positive predictivity.

References

[1] Wong D, Yip S. Machine learning classifies cancer. Nature 2018;555(7697):446–447.

[2] Hu H, Wang H, Wang F, Langley D, Avram A, Liu M. Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network. Scientific Reports 2018;8(1):4895.

[3] Safdar S, Zafar S, Zafar N, Khan NF. Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. Artificial Intelligence Review 2017;1–27.

[4] Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. The New England Journal of Medicine 2018;378(11):981.

[5] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. Journal of Machine Learning Research 2012; 13(2012):281–305.

[6] Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. In Advances in Neural Information Processing Systems. 2012; 2951–2959.

[7] Young SR, Rose DC, Karnowski TP, Lim SH, Patton RM. Optimizing deep learning hyper-parameters through an evolutionary algorithm. In Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments. 2015; 4.

[8] Rabiner LR. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 1989;77(2):257–286.

[9] Kim SJ, Kim JJ, Hahn M. Implementation and evaluation of an HMM-based korean speech synthesis system. IEICE Transactions on Information and Systems 2006;89(3):1116–1119.

[10] Kulp D, Haussler D, Reese MG, Eeckman FH. A generalized hidden markov model for the recognition of human genes in DNA. In Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology. 1996; 134–142.

[11] Altuve M, Carrault G, Beuchée A, Pladys P, Hernández AI. Online apnea–bradycardia detection based on hidden semi-markov models. Medical Biological Engineering Computing 2015;53(1):1–13.

[12] Dumont J, Hernández AI, Fleureau J, Carrault G. Modelling temporal evolution of cardiac electrophysiological features using hidden semi-markov models. In Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. 2008; 165–168.

[13] American National Standard. ANSI/AAMI EC38:1998, Ambulatory Electrocardiographs, 1998.

[14] Welch LR. Hidden markov models and the baum-welch algorithm. IEEE Information Theory Society Newsletter 2003;53(4):10–13.

[15] Murphy K. Hidden markov model (HMM) toolbox for Matlab. `https://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html`. Accessed: 2018-02-1.

[16] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine 2001;20(3):45–50.

[17] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[18] Sörnmo L, Laguna P. Bioelectrical signal processing in cardiac and neurological applications. Academic Press, 2005.

Address for correspondence:

Miguel Altuve
Km. 7 Autopista a Piedecuesta, Universidad Pontificia Bolivariana, Bucaramanga 681008, Colombia.
miguel.altuve@upb.edu.co