

Morphology Analysis of Physiological Signals Using Hidden Markov Models

D. Novák, L.Lhotská
Department of Cybernetics
Czech Technical University in Prague
xnovakd1@hpk.felk.cvut.cz

D.Cuesta-Frau, P. Micó
Department of Computer Science
Polytechnic University of Valencia, Spain
dcuesta@disca.upv.es

T.Al-ani, Y. Hamam
A2SI Laboratory
Group ESIEE-Paris, France
alanit@esiee.fr

M.Aboy
Biomedical Signal Processing Laboratory
Portland State University, USA
mateoaboy@ieee.org

Abstract

We describe a clustering algorithm based on continuous Hidden Markov Models (HMM) to automatically classify both electrocardiogram (ECG) and intracranial pressure (ICP) beats based on their morphology. The algorithm detects, classifies and labels each beat based on morphology. In order to avoid the numerical problems with classical Expectation-Maximization (EM) algorithm we apply a novel method of simulated annealing (SIM) for HMM optimization. We show that better results are achieved using simulated annealing approach.

1. Introduction

Computer-aided medical applications is a field of enormous development in recent years. One of these applications consists of extracting significant information from raw data like in case of Holter ECG signals and intracranial pressure signals.

Holter signals are ambulatory long-term ECG registers used to detect heart diseases which are difficult to find in normal electrocardiograms. These signals normally include a quantity of beats greater than 100000, and doctors must visually examine all of them in order to find possible abnormal beats. Examining every beat present in the Holter register is time consuming, and is quite likely some beats could be omitted in the visual inspection because of subjective reasons. Nevertheless, these signals include many similar beats, and only a few are different, namely, most of the time doctors are examining the same kind of beat. Therefore, it would be very useful to have a method to simplify the Holter register prior to its visual inspection.

Traumatic brain injury (TBI) remains a significant cause

of mortality and morbidity in both children and adults. TBI often leads to increased intracranial pressure (ICP) that may result in worsening brain injury and outcome. Several researchers, however, have performed preliminary studies on the ICP beat morphology and suggest that the ICP beat may contain indirect information about the intracranial compliance [5].

We introduce an automatic clustering algorithm based on continuous HMM which can be used to perform morphological analysis. In the past, cluster analysis techniques have focused on data described by static features. In many real applications, the dynamic characteristics, i.e., how a system interacts with the environment and evolves over time, are of interest. Such behavior or characteristic of these systems is best described by temporal features whose values change significantly during the observation period, like in case of electrocardiogram or intracranial pressure signals. An HMM is a very suitable tool for coping with temporal information [8].

This stochastic based modelling approach has found its use in many application areas including speech processing, molecular biology and robotics. Due to the numerical and initialisation problems we cast the problem of HMM optimization under the framework of simulated annealing [4].

2. Method

2.1 Hidden Markov Models

An HMM is a stochastic finite state automata defined by the following parameters $\lambda = (A, p, B)$, where A is a state transition probability, p is the initial state probability and B is the emission probability density function of each state is defined by a finite multivariate Gaussian mixture. Each model can be used to compute the probability of observing

a discrete input sequence $\mathbf{O} = O_1, \dots, O_T, P(\mathbf{O}|\lambda)$ to find the corresponding state sequence that maximizes the probability of the input sequence, $P(Q|\mathbf{O}, \lambda)$, and to induce the model that maximizes the probability of a given sequence $P(\mathbf{O}|\lambda') > P(\mathbf{O}|\lambda)$. The following keywords are known as the three problems of an HMM: evaluation, generation, and training.

Expectation-Maximization [6] algorithm is a standard solution for solving the training problem. However, due to the fact that classical EM algorithm is very sensitive to initialisation and also leads to meaningless parameters estimation when the EM converged to the boundary of the parameter space (where the likelihood is unbounded), we used simulated annealing approach for global HMM training. A detailed description of the procedure might be found in [8].

Simulated annealing is a well known general heuristic approach to combinatorial optimisation. Given the observation sequence \mathbf{O} , a state sequence Q is generated at random and the logarithm of probability $P(\mathbf{O}|\lambda)$ of generating \mathbf{O} is considered to be the objective value $f(Q)$ to be minimized. The solution structure is based on the choice of a state trajectory.

The various building blocks were proposed as follows [4]: (i) The initial solution is obtained simply by generating a random state trajectory. (ii) The initial temperature should be high (close to 1) to allow virtually all transitions to be accepted. Thereafter, the temperature is decreased at each iteration by a factor 0.98. (iii) The number of trials at each temperature should progressively increase with the decrease in temperature (in our case by a factor 1.02). (iv) Defining the neighbourhood structure. Moving from one solution to the next is obtained by choosing at random a state at a randomly chosen instant and affecting it randomly to another state as it is shown in Figure 1. (v) Updating the objective value. The objective function to be minimised is the overall probability of the observation sequence. Since only one state is changed at one period, cost updating is performed by calculating the probability differential. The cost evaluations of the algorithm are made independent of the problem size in order to minimize the computation time.

2.2 Clustering approach

Let's consider a data set D consisting of L sequences, $\mathcal{O} = \{\mathbf{O}_1, \dots, \mathbf{O}_L\}$. $\mathbf{O}_i = \{O_1^i, \dots, O_T^i\}$ is a sequence of length T composed of multivariate feature vectors O_t^i . The objective is to group all sequences into K clusters using the set Λ of K HMMs $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$. This problem is normally handled by methods such as k-means and Gaussian mixture model. The assumption underlying our method of clustering is that all of the sequences that belong to a cluster were generated by the same HMM and, such as, have high probabilities under this HMM. Each cluster is

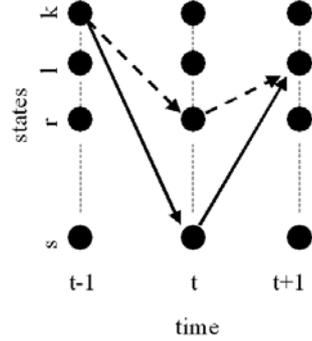


Figure 1. Defining the neighbourhood structure in simulated annealing is equaled to changing the state at period t from state s to state l .

therefore presented by one HMM model.

A natural probabilistic model for this problem is that of finite mixture model

$$P_K(\mathbf{O}) = \sum_{j=1}^K P_j(\mathbf{O}|\lambda_j)\omega_j \quad (1)$$

where \mathbf{O} denotes a sequence, w_j is the weight of the j th model; in other words it is the prior probability of component model λ_j . $P_j(\mathbf{O}|\lambda_j)$ is the density function for the sequence \mathbf{O} given a component model P_j with parameters λ_j .

To be able to perform clustering, we need some measure of sequence assignment to the clusters. The most natural way is to use the following probability (log-likelihood) that the sequence \mathbf{O}_i is generated by the HMM λ_j

$$P_{ij}(\mathbf{O}_i|\lambda_j) = \log \sum_{l=1}^N \alpha_T(l) \quad (2)$$

which rises automatically during HMM learning. Here α is the forward component of the Forward-Backward procedure (see [8]). We will further refer to this measure as *sequence-to-likelihood* measure.

Hierarchical approach, inspired by [9], is depicted by the following algorithm:

1. We train an N -states HMM for each sequence \mathbf{O}_j , ($1 \leq j \leq L$) of the training set \mathcal{O} . These L HMM are identified by λ_i and have been initialized with the k-means technique in case of EM optimization.
2. For each model λ_i , we evaluate its probability to generate the sequence \mathbf{O}_j , $1 \leq j \leq L$, obtaining a measure matrix W

$$W_{ij} = P(\mathbf{O}_j|\lambda_i), \quad 1 \leq i, j \leq L \quad (3)$$

3. We apply a complete link agglomerative hierarchical clustering to the matrix W obtaining K clusters on the data set \mathcal{O} . The algorithm produces a sequence of clustering of decreasing number of clusters at each step. The clustering produced at each step results from the previous one by merging two clusters into one.
4. Having pooled the sequences into K groups we fit each HMM to each cluster using all the observation sequences in cluster for HMM training. The weights of mixture model (1) are computed after redistribution as follows

$$\omega_j = \frac{N_j}{N} \quad (4)$$

where N_j is the number of sequences that belong to cluster j .

First, we need to symmetrize the matrix W because the result of step 2 is not a distance matrix. Thus, we define

$$W_S^{ij} = \frac{1}{2}(W_{ij} + W_{ji}) \quad (5)$$

Another kind of HMM based measure that we applied, which reminds of the Kullback-Leibler measure [7], defines the distance W_{KL} between two HMM λ_i and λ_j , and its symmetrized version W_{KLS} , as

$$W_{KL}^{ij} = W_{ii} \left(\ln \frac{W_{ii}}{W_{ji}} \right) + \left(\ln \frac{W_{ij}}{W_{jj}} \right)$$

$$W_{KLS}^{ij} = \frac{1}{2}(W_{KL}^{ij} + W_{KL}^{ji}) \quad (6)$$

Finally, we introduced another measure, called BP metric [7], defined as

$$W_{BP}^{ij} = \frac{1}{2} \left(\frac{W_{ii} - W_{ij}}{W_{ii}} + \frac{W_{jj} - W_{ji}}{W_{jj}} \right) \quad (7)$$

and motivated by the following considerations: the measure (3) defines a similarity measure between two sequences O_i and O_j as the likelihood of the sequence O_i with respect to the model λ_j , trained on O_i without really taking into account the sequence O_j . In other words, this kind of measure assumes that all sequences are modelled with the same quality without considering how well sequence O_j is modelled by the HMM λ_j .

2.3 Trace segmentation

To further alleviate computational burden we have performed trace segmentation as a feature extraction method. Trace segmentation is originally a method used to reduce the length of a discrete signal, without a great loss of information. Here, a modified approach called non-uniform sampling method, which was proposed by [2], is applied. This procedure is based on detecting changes in the signal from the amplitude of its derivative. Hence instead of using directly amplitude information we obtain after the segmentation the observation sequence, which consists of two features. The first feature is the duration of the line segment, t_i , and the second feature is the amplitude of the start point of the line segment, h_i . Therefore the observation sequence may be described as follows

$$O_i = \begin{bmatrix} t_i \\ h_i \end{bmatrix} \in \mathbb{R}^2, 1 \leq i \leq T \quad (8)$$

3 Results and Discussion

Since the classical Baum-Welch approach is a local optimization method, we used k -means initialization to increase chance of starting training procedure near to global optima. Regarding the experimental set-up, the maximum number of training iterations was $CYC_{HMM} = 100$, in case of k -means the maximum number of iterations was also $CYC_{GMM} = 100$. The convergence threshold was $\epsilon = 1e - 3$. To increase statistical significance, the experiments were repeated 10 times.

3.1 Holter ECG results

The MIT/BIH arrhythmia database was used as a gold standard [3]. The signals were denoised using wavelet based filter and the baseline signal removal has been eliminated. Then the characteristic points of ECG signals as QRS complex, P and T wave were detected and each beat was consequently isolated.

There are no apparently significant deviations in performance between different measures S, BP, KL in case of the hierarchical method as Table 1 suggests. The best results in average are achieved using W_S measure (5). In general, the average number of classes in one ECG Holter register varies between two to five, so the HMM clustering performance of 99.5% in case of five classes, W_S measure and simulated annealing is reasonable.

3.2 ICP results

We used an automatic pressure detection algorithm to detect the ICP beat components. A detailed description of the algorithm can be found in [1].

Table 1. ECG Data clustering results for three, five and seven classes. The means and variances (in parenthesis) of clustering performance are displayed. The number of HMM hidden states used was 10.

	S	KL	BP
K=3, EM	100.0(0.0)	100.0(0.0)	100.0(0.0)
K=3, SIM	100.0(0.0)	100.0(0.0)	100.0(0.0)
K=5, EM	94.5(8.4)	92.9(9.1)	93.6(9.1)
K=5, SIM	99.5(0.9)	92.9(10.8)	95.2(9.1)
K=7, EM	76.3(5.6)	78.8(4.8)	77.9(5.0)
K=7, SIM	75.7(3.0)	74.2(3.2)	73.7(3.3)

Table 2. ICP Data clustering results. The number of HMM states used were 8.

	S	KL	BP
EM	73.5(3.5)	74.5(7.1)	76.3(8.8)
SIM	75.0(9.1)	77.5(9.8)	78.6(7.5)

SIM technique outperforms the EM approach as it can be seen in Table 2. For the case of measure W_{KL} we show the symmetrized likelihood distance matrix as a color/gray-scale image in Figure 2. The axes have been ordered so that the sequences from the same clusters are adjacent. The difference in distances between the two clusters is apparent.

4 Conclusion

We applied a clustering algorithm based on HMM on both real Holter and intracranial pressure data. We implemented simulated annealing approach for HMM training. This global optimisation technique avoids well-known numerical (in several cases the EM algorithm had to be restarted) and initialisation problems of classical EM local optimisation algorithm. SIM approach is a computational intensive method. A final comparison of agglomerative hierarchical algorithm with different dissimilarity measures underlines that there are no remarkable differences among the proposed measures.

Acknowledgment

This work has been supported by the research program MSM 210000012 'Transdisciplinary Biomedical Engineer-

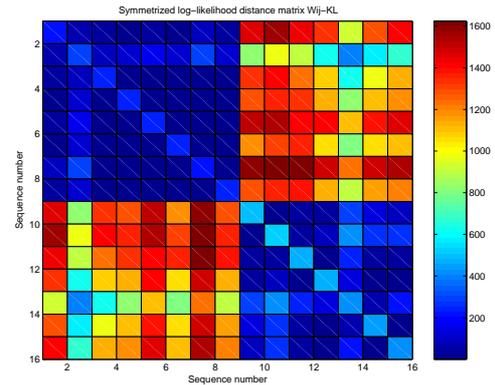


Figure 2. Example of symmetrized log-likelihood distance matrix of W_{KL} measure.

ing Research' sponsored by the Ministry of Education, Youth and Sports of the Czech Republic.

References

- [1] M. Aboy, C. Crespo, J. McNames, and B. Goldstein. Automatic detection algorithm for physiologic pressure signal components. In *Engineering in Medicine and Biology Society & Biomedical Engineering Society, 24th International Conference of the IEEE*, volume 1, pages 196–197, 2002.
- [2] D. Cuesta. Dynamic time warping applied to the clustering of electrocardiographic signals. In *European Medical and Biological Engineering Conference EMBEC99*, pages 504–505, 1999.
- [3] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. 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*, 101(23):e215–e220, 2000.
- [4] Y. Hamam and T. Al-Ani. Simulated annealing approach for training hidden markov models. In *Working Conference on Optimization-Based Computer-Aided Modeling and Design*, ESIEE, France, 1996.
- [5] G. Y. Larsen and B. Goldstein. Increased intracranial pressure. *Pediatrics in Review*, 20:234–239, 1999.
- [6] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 1997.
- [7] A. Panuccio, M. Bicego, and V. Murino. A hidden markov model-based approach to sequential data clustering. In R. D. M. K. T. Caelli, A. Amin and D. de Ridder Eds, editors, *Structural, Syntactic and Statistical Pattern Recognition, LNCS 2396*, pages 734–742. Springer, 2002.
- [8] R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77, 1989.
- [9] P. Smyth. Clustering sequences with hidden markov models. *Advances in Neural Information Processing Systems*, (9), 1997.