JOINT SEGMENTATION AND AR MODELING OF QUASISTATIONARY SIGNALS USING THE EM ALGORITHM

Ignacio Santamaría-Caballero, Carlos Pantaleón-Prieto, Aníbal R. Figueiras-Vidal *

Dpto. Ing. Comunicaciones, ETSII y Telecom. University of Cantabria
Avda. Los Castros, 39005 Santander, Spain
* EPS, Universidad Carlos III de Madrid , Butarque 15, 28911 Leganés (Madrid), Spain
e-mail: nacho@gtas.dicom.unican.es

ABSTRACT

This paper presents a method for off-line segmentation and AR modeling of signals characterized by abrupt changes between stationary segments (quasistationary signals). Assuming that the number of models and their orders are known, we propose a suboptimal procedure for maximizing the likelihood function based on the Expectation-Maximization algorithm. At each iteration the transitions are estimated as the posterior probabilities that a sample was generated by a given model (E-step); then, the new set of models is obtained by solving a least squares problem (M-step). It is shown by means of computer simulations that the algorithm achieves accurate estimates of the transitions and AR coefficients with a moderate computational complexity.

1. INTRODUCTION

Autoregressive (AR) modeling is a well-known technique for stationary signal analysis which has found application in areas such as speech coding and spectral estimation. Signals encountered in practice, however, are usually non stationary; a frequent class of them, referred to as quasistationary or locally stationary [1], is characterized by abrupt changes between stationary segments with different statistical properties. In these situations, the performance obtained by a single (global) AR model can be improved by using a different (local) AR model for each stationary segment.

A complete solution to this problem consists of finding the correct number of models (or stationary segments), their orders and parameters, and the transitions among them. A Bayesian solution has been proposed in [2], where the *a posteriori* probability distribution is maximized via dynamic programming; however, this approach has a very high computational cost. An alternative solution has been given in [3,4], where it is shown that if different linear models compete using a "winner takes all" strategy (i.e., each sample is assigned to the model producing the smaller prediction error), each model concentrates on a stationary time segment of the signal. After competition, the obtained models can be used to estimate the boundaries between stationary segments.

In this paper we consider a simpler version of the problem, where the number of models and their orders are known: this represents a typical situation, for instance, in speech modeling, where it is interesting to evaluate if there is some improvement when a single AR model of a given length is substituted by two shorter AR models and a segmentation law (keeping fixed the total number of parameters). To find a solution in this simpler case is still a difficult problem, since it is necessary to perform a search over a highly multidimensional space composed of all the possible transitions and AR parameters.

We propose an off-line maximum likelihood solution, which allows the simultaneous segmentation of the quasistationary time series and the estimation of the optimal AR models. The maximization of the likelihood function is carried out using the Expectation-Maximization (EM) algorithm [5]: at each step, the transitions are estimated by computing the conditional probabilities that a particular sample was generated by a given model; then, a new set of models is obtained solving a least squares problem.

2. PROBLEM STATEMENT

2.1. Signal model

We consider that the observations are generated by switching among M different AR models of orders $p_1, \cdots, p_M$, and coefficients $\mathbf{a}_j = (a_{1,j}, \cdots, a_{p_j,j})$; i.e.,

This work has been supported by CICYT grant TIC96-0500-C10-07

$$x[n] = \sum_{j=1}^{M} t_j[n] x_j[n] \tag{1}$$

where $t_j[n]$ selects the samples generated by model j

$$t_j[n] = \begin{cases} 1, & \text{if } x[n] \text{ is generated by model } j \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

On the other hand, the output at time instant n for model j is given by

$$x_j[n] = -\sum_{i=1}^{p_j} a_{i,j} x[n-i] + e_j[n] \qquad j = 1, \cdots, M \tag{3}$$

where $e_j[n]$ is a zero mean uncorrelated Gaussian noise with variance σ_j^2.

Note that (3) is generated using p_j past samples of the observations: when a transition occurs these past values are used as the initial condition for the new model after change. This method of generating a quasistationary signal results in a smooth transition between the observations before and after the change. This process is depicted in Fig. 1, where $A_j(z) = 1 + \sum_{i=1}^{p_j} a_{i,j} z^{-i}$.

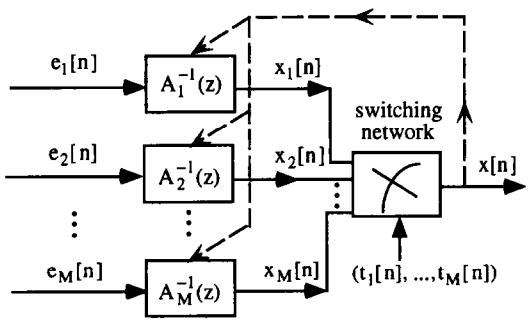

Figure 1: Method of generating a quasistationary signal.

Now, the problem can be stated as follows: given the number of models M, their orders $(p_1, \cdots, p_M)$, and the vector of observations $\mathbf{x} = (x[0], \cdots, x[N-1])^T$, first, determine the boundaries between segments; and second, find the best model for each segment.

2.2. Cost function

Our objective in this section is to find the likelihood function for the unknown parameters. For notation purposes, let us group the coefficients of the AR models, the variances of the noise sequences and the transition sequences in vectors $\mathbf{a} = (\mathbf{a}_1, \cdots, \mathbf{a}_M)^T$, $\sigma = (\sigma_1^2, \cdots, \sigma_M^2)^T$ and $\mathbf{t} = (\mathbf{t}_1, \cdots, \mathbf{t}_M)$; respectively.

Using model (1), the density function for each sample $x[n]$ is given by

$$p(x[n]; \mathbf{t}, \mathbf{a}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{ -\frac{1}{2\sigma_n^2} \left(x[n] + \sum_{j=1}^{M} t_j[n] \sum_{i=1}^{p_j} a_{i,j} x[n-i] \right)^2 \right\} \tag{4}$$

where $\sigma_n^2 = \sum_{j=1}^{M} t_j[n] \sigma_j^2$.

Assuming independence among the stationary segments, we have that the log likelihood function for the unknown parameters is given by

$$L(\mathbf{x}; \mathbf{t}, \mathbf{a}, \sigma) = -\sum_{n=\max(p_j)}^{N-1} \frac{1}{2\sigma_n^2} \left(x[n] + \sum_{j=1}^{M} t_j[n] \sum_{i=1}^{p_j} a_{i,j} x[n-i] \right)^2 - \sum_{n=\max(p_j)}^{N-1} \log \sigma_n. \tag{5}$$

Since the evaluation of (4) is only possible for $n > \max(p_j)$, the summation in (5) runs from $n = \max(p_j)$ to $N-1$.

3. PROPOSED SOLUTION

3.1. Applying the EM algorithm

A direct maximization of (5) cannot be carried out in practice due to the discrete nature of the transition sequences $\mathbf{t}$. To overcome this difficulty we apply a suboptimal procedure based on the EM algorithm [5], which increases the likelihood of the obtained estimates, iteration by iteration, until a local maximum is reached.

The observed incomplete data is $\mathbf{x}$; the unobserved data is the vector of transition sequences $\mathbf{t}$. Using this choice for the complete data set $(\mathbf{x}, \mathbf{t})$, and denoting the current estimates of the parameters after k iterations of the EM algorithm as $\hat{\mathbf{t}}_k$, $\hat{\mathbf{a}}_k$ and $\hat{\sigma}_k$; then, the next iteration cycle is given by

$$\begin{aligned} \textit{Estep} &: \hat{\mathbf{t}}_{k+1} = \mathrm{E}[\mathbf{t} | \mathbf{x}, \hat{\mathbf{a}}_k, \hat{\sigma}_k] \\ \textit{Mstep} &: (\hat{\mathbf{a}}_{k+1}, \hat{\sigma}_{k+1}) = \operatorname{argmax} L(\mathbf{x}; \hat{\mathbf{t}}_{k+1}, \mathbf{a}, \sigma) \end{aligned} \tag{6}$$

where $\mathrm{E}[\cdot]$ denotes expectation.

First, let us consider the E-step: $t_j[n]$ can be estimated as the conditional (posterior) probability of model j given $x[n]$, i.e.,

$$\hat{t}_j[n] = p(Model_j|x[n]). \quad (7)$$

In (7) the subscript denoting iteration is understood. Now, applying Bayes, and assuming equal *a priori* probabilities for each model, (7) can be rewritten as

$$\hat{t}_j[n] = \frac{p(x[n]|Model_j)}{\sum_{k=1}^{M} p(x[n]|Model_k)}. \quad (8)$$

Finally, the probabilities $p(x[n]|Model_j)$ can be obtained from (4) considering only the prediction error for model j

$$p(x[n]|Model_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp - \left(\frac{e_j^2[n]}{2\sigma_j^2} \right) \quad (9)$$

In fact, the transition estimates can be viewed as soft decision functions similar to those proposed in the neural network literature to design modular networks [6] or in unsupervised competitive learning algorithms [7]: this soft scheme improves the competitive local linear procedure proposed in [3], where a hard decision technique is used.

Once the transition sequences have been obtained, the M-step consists of estimating the parameters of the AR models: coefficients and variances. A standard procedure to get the AR coefficients consists of solving the following linear least squares problem

$$x[n] = -\sum_{j=1}^{M} \hat{t}_j[n] \left(\sum_{i=1}^{p_j} a_{i,j} x[n-i] \right) \quad (10)$$

for $n = \max(p_j), \cdots, N-1$; which can be rewritten in matrix notation as

$$[\mathbf{T}_1\mathbf{X}_1|\cdots|\mathbf{T}_M\mathbf{X}_M]\,\mathbf{a} = \mathbf{x} \quad (11)$$

where $\mathbf{X}_j$ is a $(N - \max(p_j)) \times p_j$ matrix which rows are delayed observations vectors (the rth row is $(x[r-1], \cdots, x[r-p_j])$, and $\mathbf{T}_j$ are $(N - \max(p_j)) \times (N - \max(p_j))$ diagonal matrices formed by the transition sequences estimated in the E-step.

From (10), we can see that to predict $x[n]$, each model contributes in a proportion given by the probabilities $\hat{t}_j[n]$. If over time a sample is better predicted by a particular model, its posterior probability is pulled towards 1; otherwise it approaches 0.

Finally, the variances of each model are estimated according to

$$\sigma_j^2 = \frac{\sum_n e_j^2[n] t_j[n]}{\sum_n t_j[n]}, \quad (12)$$

The direct procedure proposed in this section has the drawback that isolated samples within a stationary segment could be better predicted by a model corresponding to a different segment. To avoid this undesirable effect which causes instantaneous transitions among the models, we propose to evaluate (7) using not only the sample to be predicted, but a neighborhood around it. Specifically, considering a window of length $2l+1$ centered in $x[n]$, and assuming that all the samples belongs to the same model, Eq. (9) can be substituted by

$$p(x[n-l], \cdots, x[n], x[n+l]|\mathbf{a}_j, \sigma_j) = \prod_{k=-l}^{k=l} \frac{1}{\sqrt{2\pi}\sigma_j} \exp - \left(\frac{e_j^2[n-k]}{2\sigma_j^2} \right). \quad (13)$$

The length of the window allows a tradeoff between the resolution to detect the boundaries between segments and the ability to avoid instantaneous transitions.

3.2. Initialization via competitive modeling

The log likelihood function (5) is rife with local maxima, therefore a proper initialization of the EM algorithm is a key factor in obtaining a satisfactory solution.

A simple alternative consists of partition the data into M nonoverlapping segments (M being the number of models), which boundaries are selected at random. The AR coefficients and variances obtained for each segment are used as initial values for the EM algorithm.

An improved initialization strategy consists of using the competitive procedure described in [3]. Starting from random coefficient AR models (or using the procedure described in the previous paragraph), we iteratively select at random a sample of the signal and choose the model that better predicts that sample. Only that model is trained, leaving the rest of the models intact, i.e., we use a "winner takes all" strategy. The coefficients of the winner model are updated using an LMS algorithm. The iterations are carried out until a previously specified error criterion is fulfilled or a maximum number of iterations is reached. As it is shown in [3], after competition each model concentrates on a different stationary segment of the signal.

Starting the EM algorithm from these tentative models clearly improves the obtained results.

4. SIMULATION RESULTS

To verify the proposed algorithm and test its performance, we conducted several Monte-Carlo experiments

with simulated data. Here we present the results obtained using the same quasistationary process considered in [2]. In particular, we generated a quasistationary process of 300 points composed of three segments of AR processes with two transitions at 81 and 211. The first process was of second order with parameters $a_{1,1} = -1.37$ and $a_{2,1} = 0.56$. The second one was of fourth order with parameters $a_{1,2} = -1.6$, $a_{2,2} = 1.73$, $a_{3,2} = -0.924$, and $a_{4,2} = 0.3816$; finally, the third one was of first order with parameter $a_{1,3} = 0.8$. The variance of the excitation noise was 1 for the three models. A typical realization of this process is shown in Fig. 2.

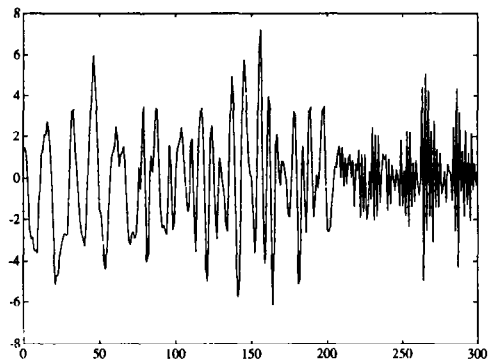

Figure 2: A typical realization of the quasistationary process.

The results summarized below were obtained from 1000 Monte-Carlo simulations. The initial values for the EM algorithm were obtained using the competition procedure described in Section 3.2. Specifically, we applied 2500 iterations with a learning rate of 0.005. With these starting values the method gave two transitions in 81 % of the cases. On the other hand, without using the competitive procedure for obtaining the initial models, this percentage reduces to 51 %.

In order to obtain the posterior probabilities for each model we used a window of length 10 (see Eq. (13)). As an example of the behavior of the proposed algorithm Figs. 3 and 4 show the posterior probabilities for model 2 at iterations 1 and 7 (after convergence), respectively. A number of iterations between 3 and 20 leads to convergence in most of the cases. The evolution of the log likelihood function for this example is shown in Fig. 5. Finally, the mean values and variances of the transition and the AR coefficient estimates are shown in Tables 1 and 2, respectively.

Comparing the obtained results with those given in [2] and [4], we can conclude that the proposed al-

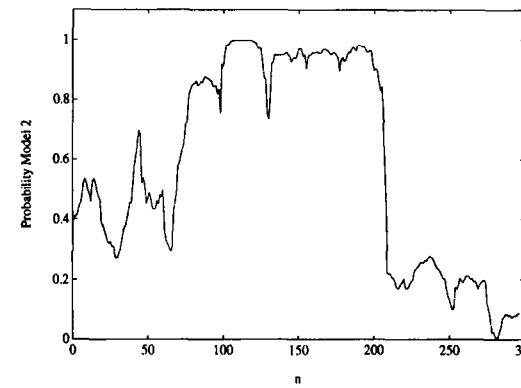

Figure 3: Posterior probabilities for model 2 ($t_2[n]$) at iteration 1.

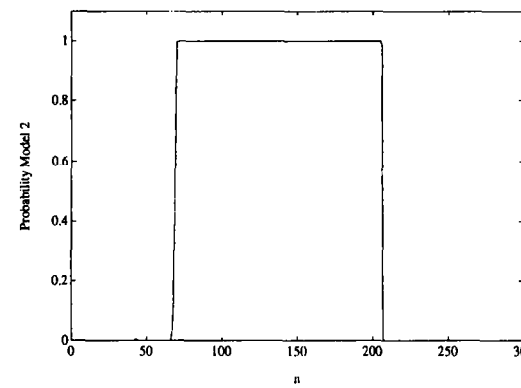

Figure 4: Posterior probabilities for model 2 ($t_2[n]$) at iteration 7.

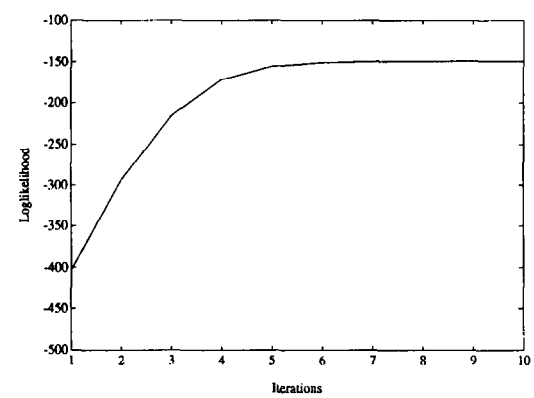

Figure 5: Evolution of the log likelihood function.

Transitions	Mean value	Standard Deviation
T1=81	79.1	7
T2=211	205.4	4.5

Table 1: Mean values and standard deviations of the transition estimates

AR coefficients	Mean value	Variance
$a_{1,1} = -1.37$	-1.349	0.013
$a_{2,1} = 0.56$	0.553	0.014
$a_{1,2} = -1.6$	-1.580	0.015
$a_{2,2} = 1.73$	1.676	0.052
$a_{3,2} = -0.924$	-0.877	0.047
$a_{4,2} = 0.3816$	0.356	0.013
$a_{1,3} = 0.8$	0.654	0.027

Table 2: Mean values and variances of the estimates of the AR coefficients

gorithm approaches the performance of the Bayesian formulation of [2] without its high computational cost and refines the results of the competitive method in [4] with a moderate increase of complexity.

5. CONCLUSIONS

This paper has presented a new method for joint segmentation and AR modeling of quasistationary signals. In particular, assuming that the number of models (or stationary segments) and their orders is known, the addressed problem consists of estimating the boundaries between segments and finding the best model for each segment.

The maximization of the likelihood function is carried out using the EM algorithm. The prediction errors obtained for a given set of models are used to estimate the probabilities that a sample was generated by a particular model (E-step) and, using these estimates, a new set of optimal models is obtained (M-step). An adequate choice of the initial values obtained via competitive modeling, contribute to achieve fast convergence and improved estimates. On the other hand, its moderate computational complexity makes the proposed algorithm a useful alternative in many signal processing applications such as vibration monitoring of mechanical systems, fault detection, time series prediction or speech modeling. The extension of the proposed approach for nonlinear models,(instead of AR), seems to us an interesting direction of future research.

6. REFERENCES

[1] M. Basseville, I. V. Nikiforov, *Detection of Abrupt Changes* , Prentice Hall, Englewood Cliffs, NJ, 1993.

[2] P. M. Djuric, S. M. Kay, G. F. Boudreaux-Bartels, "Segmentation of nonstationary signals", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP'92)*, vol. 5, pp. 161-164, 1992.

[3] C. J. Pantaleón-Prieto, I. Santamaría-Caballero, A. R. Figueiras-Vidal, "Competitive local linear modeling", *Signal Processing*, vol. 49, pp. 73-83, 1996.

[4] C. J. Pantaleón-Prieto, A. R. Figueiras-Vidal, "Nonlinear time series modeling by competitive segmentation of state space", *Proc. of the International Workshop on Artificial Neural Networks*, pp. 525-530, Sitges, Spain, 1993.

[5] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. Royal Statist. Soc.*, vol. B-39, pp. 1-37, 1977.

[6] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, G. E. Hinton "Adaptive mixtures of local experts", *Neural Computation*, vol. 3, pp. 79-87, 1991.

[7] S. J. Nowlan, "Maximum likelihood competitive learning", in D. S. Touretzky, Ed., *Advances in Neural Information Processing Systems II*, pp. 574-582; San Mateo, CA: Morgan Kaufmann; 1990.