MODELING CHAOTIC SYSTEMS WITH HIDDEN MARKOV MODELS

Cory Myers[1] *Andrew Singer*[2] *Frances Shin*[1] *Eugene Church*[3]

[1]Lockheed Sanders, Inc., PO Box 868, Nashua, NH 03061
[2]Research Laboratory of Electronics, MIT, Cambridge, MA 02139
[3]ARDEC, Picatinny, NJ 07801

ABSTRACT

The problem of modeling chaotic nonlinear dynamical systems using hidden Markov models is considered. A hidden Markov model for a class of chaotic systems is developed from noise-free observations of the output of that system. A combination of vector quantization and the Baum-Welch algorithm is used for training. The importance of this combined iterative approach is demonstrated. The model is then used for signal separation and signal detection problems. The difference between maximum likelihood signal estimation and maximum aposteriori signal estimation using a hidden Markov model is illustrated for a nonlinear dynamical system.

1. INTRODUCTION

Recent work in the areas of nonlinear dynamics and chaos has demonstrated that "noise-like" signals can be produced by nonlinear dynamical systems [1]. These signals appear to be broadband and random, yet their behavior is deterministic. Not only have chaotic signals been observed to be produced by real systems, but nonlinear dynamical models have also shown promise in many signal processing problems [2, 3].

As a result of research aimed at explaining and exploiting the behavior of chaotic systems, a number of new signal processing problems have emerged. While initial attention had focused on the study of the richness of the behavior and properties of chaotic systems, there is now considerable interest in the problem of modeling data based on chaotic systems and in addressing problems of signal processing for such systems.

1.1. Chaotic Dynamical Systems

We assume that a chaotic dynamical system is modeled by a nonlinear state equation

$$s_{n+1} = f(s_n) \qquad (1)$$

where $s_n \in R^m$ is the state of the system at time index n and f is a nonlinear function. A nonlinear dynamical system operating in a chaotic regime is characterized by the presence of an *attractor*. Qualitatively, an attractor is the portion of state space such that, after an initial transient, the nonlinear dynamical system settles down to. The presence of an attractor implies that the system, in a sense, repeatedly visits some portion of state space, without necessarily being periodic. A more complete description of the properties of chaotic systems and their physical importance can be found in [4].

This work sponsored in part by the US Army under contract DAAL03-86-D-0001, Delivery Order 2293, in part by the Office of Naval Research under contract N00014-91-C-0125, and in part by the Army Research Office under contract DAAL03-91-C-0052.

1.2. The System Modeling Problem

We assume that we observe the state of a chaotic system through an observation function, h: $y_n = h(s_n)$, where $y_n \in R^k$ is the output from the chaotic system. We have used one-dimensional observations in all of our work to date. The system modeling problem is, given a sequence of observations $y_{1:N} = \{y_1, y_2, \ldots, y_N\}$, to develop a model for the system that can be used for analysis of the system, prediction of future behavior of the system, signal separation, system identification, or signal detection. Depending on the modeling problem, it may not be necessary to generate an explicit model for the state update function f.

Past work in the area of system modeling for chaotic systems has considered explicit model development, prediction, and noise removal. Modeling chaotic systems from observations has focused on developing global nonlinear maps or locally linear or nonlinear maps that describe the evolution of the system [5]. Prediction is performed by extrapolating from the current state according to the map. Noise removal methods have been developed based on iterative techniques and on optimization criteria [5, 6].

2. MODELING DYNAMICAL SYSTEMS WITH HIDDEN MARKOV MODELS

By their very nature, dynamical systems are Markov processes and the presence of an attractor in a chaotic system imposes a natural probabilistic measure on the state space - the invariant density. Marteau and Abarbanel have developed an iterative noise removal procedure in which training data is used to estimate the invariant density [7].

We propose to model a nonlinear dynamical system using a *discrete state space model with continuous observations* in which we do not assume that our states necessarily have any relationship to the true states of the system. We choose our representation such that the observations are well-modeled. Thus, our model is a hidden Markov model (HMM) because the states are not observed directly. This approach is based on ideas from Fraser, who demonstrated that a HMM can capture some of the aspects of a chaotic system [8].

Our model for a nonlinear dynamical system consists of the following parts:

- A set of L states $S = \{S_1, S_2, \ldots, S_L\}$. We denote the state at time index n as q_n where $q_n \in S$.

0-7803-0532-9/92 $3.00 © 1992 IEEE

- An L by L state transition matrix A that defines the probability of the next state given the current state. We assume that the states are fully interconnected and allow the training procedure to determine the proper connectivity.
- An L element vector $\underline{\pi}$ that defines the initial state probabilities. For a chaotic system operating on an attractor *the initial state probabilities are the state probabilities regardless of time index, i.e.*, the invariant density. Equivalently, we do not assume that the initial condition for the training data is the initial condition for all outputs of the chaotic system.
- A set of L observation probability densities, one for each state, that determine the likelihood of the observations given the model. In our modeling work we have assumed that the observation density is Gaussian with a state-dependent mean $m(q)$ and variance $\sigma^2(q)$. This observation density is particularly useful because it can be easily modified to include Gaussian observation noise. We denote the $2L$ element vector of mean and variance parameters as $\underline{b}$.

Training of our model begins with a sequence of clean observations $y_{1:N} = \{y_1, y_2, \ldots, y_N\}$ and attempts to determine A, $\underline{\pi}$, and $\underline{b}$ to maximize the likelihood of the observations given the model parameters, *i.e.*,

$$\{\hat{A}, \hat{\underline{b}}, \hat{\underline{\pi}}\} = \underset{A, \underline{b}, \underline{\pi}}{\operatorname{argmax}} \Pr(y_{1:N} | A, \underline{b}, \underline{\pi}), \qquad (2)$$

where we use the Markov property:

$$\begin{aligned} &\Pr(y_{1:N}|A, \underline{b}, \underline{\pi}) \\ &= \sum_{q_{1:N}} \Pr(y_{1:N}| q_{1:N}, A, \underline{b}, \underline{\pi}) \Pr(q_{1:N}| A, \underline{b}, \underline{\pi}) \\ &= \sum_{q_{1:N}} \underline{\pi}(q_1) \Pr(y_1| q_1, \underline{b}) \prod_{n=2}^{N} \Pr(q_n| q_{n-1}, A) \Pr(y_n| s_n, \underline{b}) \end{aligned} \qquad (3)$$

in the optimization. The sums in Eq. (3) are over all possible state sequences.

We use the segmental K-means and the Baum-Welch algorithms to estimate our model parameters, as follows:

1. Create vectors from the observations using delays, *i.e.*, form the vectors $[y_n \; y_{n-1} \; \cdots \; y_{n-D+1}]$. See [4] for a discussion of the selection of an embedding dimension D.
2. Develop a vector quantizer with L codewords for the vectors. Quantize the observations using this quantizer.
3. Estimate the transition probabilities using the observed transitions in the quantized observations. Estimate the observation means as the codewords and the observation variances according to the quantization error.
4. Iteratively perform Viterbi decoding to estimate the best quantized sequence and re-estimate the model parameters using the quantized sequence.
5. Use the Baum-Welch algorithm to finalize the model parameters [9].

While we do not require that the states in our hidden Markov model have any relationship with the state of the nonlinear dynamical system, we do initialize our search for a good model by effectively quantizing state space. Figure 1 shows the two-dimensional embedding of the Henon map:

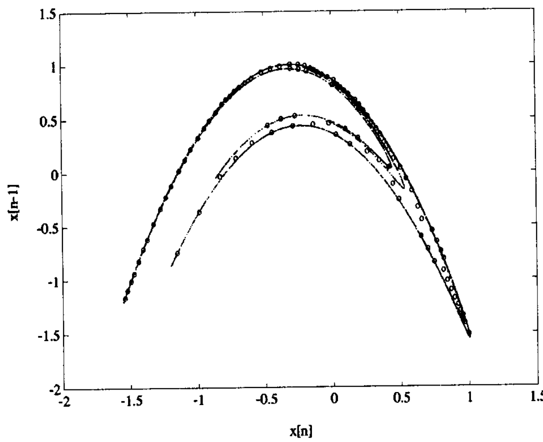

FIGURE 1. Time-delay embedding and state locations for the Henon map after vector quantization. 10,000 samples of data were used to train a 100 state model.

Model	Training Set Size	States	Non-zero Transitions	Average LL
I	7500	75	222	1.88
II	10000	100	298	2.06
III	25000	250	707	2.71

TABLE 1. Parameters of different models of the Henon map.

$$\begin{aligned} s_{1,n+1} &= 1 + s_{2,n} - 1.4 s_{1,n}^2 \\ s_{2,n+1} &= 0.3 s_{1,n} \end{aligned} \qquad (4)$$

when $h(s) = s_1$, along with the locations of the vector quantizer codewords. The state locations after vector quantization appear to be a reasonable representation of the embedded vectors. However, the likelihood function is increased by 13% by use of the Baum-Welch algorithm.

Table 1 shows the results of different modeling trials for the Henon map. The number of non-zero transitions shows the number of transitions that are present after Step 4 of the training. The results show that, on average, there are three transitions from each state. The average log likelihood is $N^{-1}\log(\Pr(y_{1:N}| A, \underline{b}, \underline{\pi}))$ and shows that, as expected, increasing the number of states increases the likelihood.

3. SIGNAL SEPARATION

A common signal processing problem is to separate a received signal into two components. We have examined this problem for chaotic systems using hidden Markov models. We assume that we are given a sequence of observations, $O_{1:N} = \{O_1, O_2, \ldots, O_N\}$, where each observation is the sum of the output of the chaotic system and another signal, *i.e.*, $O_n = y_n + w_n$, and we wish to generate the best estimates for y_n and w_n given the observations.

For our initial work, we assume that the other signal, w_n, can be modeled as white Gaussian noise with variance σ_w^2. We define two signal estimation algorithms – one based on a maximum likelihood state sequence estimation approach and one based on a maximum aposteriori approach.

3.1. Maximum Likelihood Approach

The maximum likelihood signal estimation approach first estimates the most likely state sequence given the observations, *i.e.*,

$$\hat{q}_{1:N} = \underset{q_{1:N}}{\text{argmax}}\ \Pr(q_{1:N}|\,O_{1:N}) \tag{5}$$

This is computed using the Viterbi algorithm. The signal y_n is then estimated as the expected value of y_n given the observations and the most likely state sequence, *i.e.*,

$$\begin{aligned}\hat{y}_n &= \mathrm{E}\,[y_n|\,O_n, \hat{q}_n] \\ &= m(\hat{q}_n) + \frac{\sigma^2(\hat{q}_n)}{\sigma^2(\hat{q}_n) + \sigma_w^2}(O_n - m(\hat{q}_n))\end{aligned} \tag{6}$$

where $m(\hat{q}_n)$ and $\sigma^2(\hat{q}_n)$ are the mean and variance of the most likely state at time index n.

3.2. Maximum Aposteriori Approach

In the maximum aposteriori approach we attempt to estimate the signal y_n as the expected value of y_n given the observations, *i.e.*,

$$\begin{aligned}\hat{y}_n &= \mathrm{E}\,[y_n|\,O_{1:N}] \\ &= \sum_{q_{1:N}} \mathrm{E}\,[y_n|\,O_n, q_n]\Pr(q_{1:N}|\,O_{1:N})\end{aligned} \tag{7}$$

where the summation is performed over all possible state sequences, $q_{1:N}$, and where $\mathrm{E}\,[y_n|\,O_n, q_n]$ is computed according to Eq. (6). Note that Eq. (7) can be computed efficiently using the forward-backward algorithm to compute $\Pr(q_n|\,O_{1:N})$ [9].

We note that in the case of a linear dynamical system, driven by white Gaussian noise, both the maximum likelihood and the maximum aposteriori approaches converge to Kalman smoothing as the number of states goes to infinity.

3.3. Processing Results

Figure 2 shows an example of noise removal using hidden Markov models for a chaotic system. In this example 25,000 samples from the Henon map were used to train a 250 state model. An additional 2000 data points from the Henon map were generated using different initial conditions. This is the signal y_n. White Gaussian noise was added to y_n for a signal to noise ratio of 10 dB. This signal was then processed by the maximum aposteriori processing method. The time-delay embedding of the original signal is shown in Figure 1. Figure 2a shows the time-delay embedding for the noisy signal. Figure 2b shows the time-delay embedding for the output of maximum aposteriori cleaning. The signal to noise ratio in the output is 21 dB, a processing gain of 11 dB. Table 2 compares the processing gain of the maximum likelihood and the maximum aposteriori methods for different size models (see Table 1), both with and without the Baum-Welch iteration. The results show improving performance with increasing model size at the higher signal to noise ratios. Also shown is the improvement gained by using a model built with the Baum-Welch iteration and in using the maximum aposteriori method. Table 3 compares the maximum aposteriori method with the scaled probabilistic method of Marteau and Abarbanel and with Wiener filtering. These results show that the maximum aposteriori processor is 4 to 8 dB better than the scaled probabilistic method and 4 to 12 dB better than Weiner filtering for the Henon data.

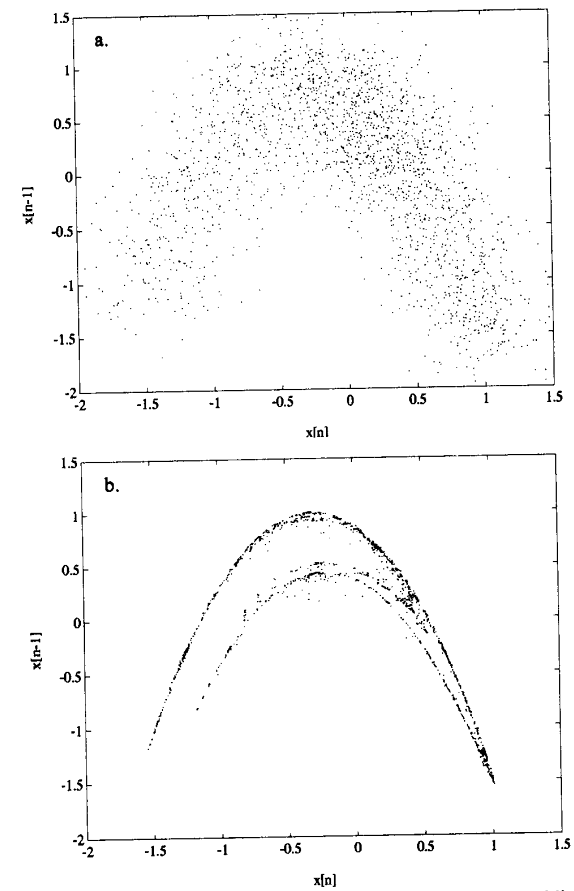

FIGURE 2. Noise removal example. Time-delay embedding of noisy signal. b. Time-delay embedding of recovered signal.

Model	ML or MAP	Baum-Welch	Input SNR 0 dB	10 dB	20 dB
I	MAP	yes	6.9	19.9	26.8
II	MAP	yes	6.5	19.4	28.3
III	MAP	yes	6.9	21.5	32.8
III	MAP	no	6.4	20.2	27.5
III	ML	yes	4.2	20.0	31.4

TABLE 2. Noise removal on the Henon map – output SNR (dB) for various input SNRs and test conditions.

Figure 3 shows an example of recovery of a signal buried in chaotic noise via a hidden Markov model. A recording of a helicopter signal was added to the output of the Henon map. The ratio of the helicopter signal to the chaotic signal was -20 dB. Figure 3a shows the original helicopter signal. Figure 3b shows the sum of the helicopter signal and the chaotic noise. Figure 3c shows the result of maximum aposteriori processing. The resulting signal to noise ratio is 7 dB.

Method	Input SNR 0 dB	10 dB	20 dB
MAP using HMM	6.9	21.5	32.8
Scaled Probabilistic	2.3	13.6	24.4
Weiner Filtering	3.0	10.4	20.0

TABLE 3. Comparison of noise removal based on HMM, scaled probabilistic method, and Weiner filtering.

4. SIGNAL DETECTION

A model for the generation of signals is not only useful for signal separation, but it is also useful for signal detection. We have experimented with using a hidden Markov model for the detection of chaotic signals. Specifically, we train a model using a clean version of the chaotic signal and then attempt to detect the chaotic signal in background noise.

We use the log likelihood ratio as our detection statistic, z:

$$z = \log(\Pr(O_{1:N}|A, \underline{b}, \underline{\pi})) - \log(\Pr(O_{1:N}|H_0)), \quad (8)$$

where the first term is the probability that the observations are the sum of a chaotic signal plus Gaussian noise and the second term is the probability that the observations are Gaussian noise. The first term is calculated using the forward-backward algorithm.

Table 4 shows the results of signal detection for the Henon map using a hidden Markov model containing 250 states. In each case the detector is presented 100 samples of a signal containing the output of the Henon map in noise. The performance is evaluated in two ways: first a "pseudo SNR" is calculated as

$$\text{"SNR"} = \frac{(E[z|\text{ signal present}] - E[z|\text{ noise only}])^2}{\text{Var}[z|\text{ noise only}]}, \quad (9)$$

secondly, we compute, via Monte Carlo simulation, the probability of detection at various false alarms rates. As a comparison, we show in Table 5 the expected performance of a noncoherent integrator at the same signal to noise ratios assuming 100 data points. The results show that a detector based on a hidden Markov model performs significantly better than a noncoherent detector. However, the performance does not approach that of a coherent detector. For example, at a −10 dB input SNR with 100 samples, a coherent detector will achieve a P_D of 0.94 at a P_{FA} of 0.05.

5. SUMMARY AND FUTURE WORK

We have presented initial results on modeling chaotic dynamical system using hidden Markov models. We have shown that a hidden Markov model is useful for signal cleaning and for signal detection of chaotic systems. Currently we are working to combine such modeling techniques with other methods for modeling chaotic dynamical systems, to develop processing techniques for multi-dimensional data, and to apply our work to training from noisy observations.

REFERENCES

[1] T. S. Parker and L. O. Chua, "Chaos: A Tutorial for Engineers," *Proceedings of the IEEE*, Vol. 75, No. 8, August 1987.

[2] T. F. Quatieri and E. M. Hofstetter, "Short-Time Signal Representation for Nonlinear Difference Equations," *Proceedings of ICASSP 1990*, pp. 1551-1554.

[3] A. C. Signer, G. W. Wornell and A. V. Oppenheim, "Codebook Prediction: A Nonlinear Signal Modeling Paradigm," to appear in *Proceedings of ICASSP 1992*.

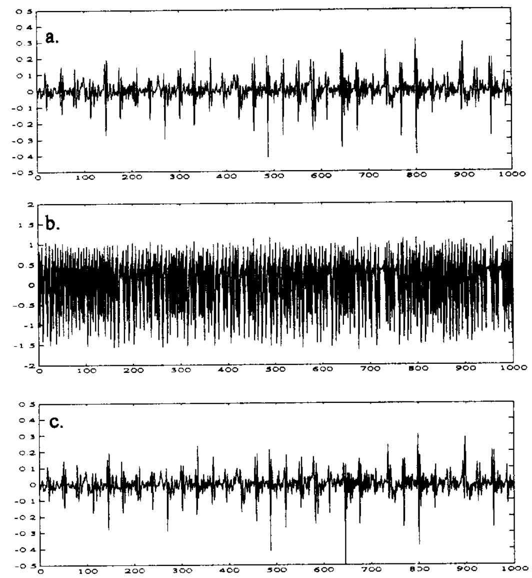

FIGURE 3. Signal recovery example. a. Original helicopter signal. b. Noisy signal. c. Recovered signal.

Input SNR	Output "SNR"	P_D: P_{FA} = .01	P_{FA} = .05	P_{FA} = .10
-20	-9.6	.04	.14	.20
-15	-1.7	.10	.26	.36
-10	1.2	.08	.26	.44
-5	11.3	.79	.92	.97

TABLE 4. Results for detector based on hidden Markov model.

SNR	P_{FA} = .01	P_{FA} = .05	P_{FA} = .10
-20	.01	.06	.12
-15	.02	.08	.15
-10	.06	.18	.29
-5	.40	.64	.76

TABLE 5. Probability of detection for a noncoherent detector using 100 input samples.

[4] H. D. I. Abarbanel, "Chaotic Signals and Physical Systems, to appear in *Proceedings of ICASSP 1992*.

[5] J. D. Farmer and J. J. Sidorowich, "Optimal Shadowing and Noise Reduction," *Physica D*, No. 47, pp. 373-392, 1991.

[6] E. J. Kostelich and J. A. Yorke, "Noise Reduction: Finding the Simplest Dynamical System Consistent with the Data," *Physica D*, No. 41, pp. 183-196, 1990.

[7] P.-F. Marteau and H. D. I. Abarbanel, "Noise Reduction in Chaotic Time Series Using Scaled Probabilistic Methods," to appear in *Journal of Nonlinear Science*.

[8] A. M. Fraser, "Chaotic Data and Model Building," *NATO ASI Proceedings of Information Dynamics*, 1991.

[9] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257-286, February 1989.