

Sampling Theory for Process Detection with Applications to Surveillance and Tracking

Diego Hernando^a and Valentino Crespi^b

^aThayer School of Engineering at Dartmouth College, 8000 Cummings Hall, Hanover, NH;

^bCalifornia State University, Los Angeles, 5151 State University Drive, Los Angeles, CA

ABSTRACT

In this paper, we investigate the link between the rate at which events are observed by a monitoring system and the ability of the system to effectively perform its tracking and surveillance tasks. In general, higher sampling rates provide better performance, but they also require more resources, both computationally and from the sensing infrastructure.

We have used Hidden Markov Models to describe the dynamic processes to be monitored and (α, β) -currency as a performance measure for the monitoring system. Our ultimate goal is to be able to determine the minimum sampling rate at which we can still fulfill the performance requirements of our system.

Along with the theoretical work, we have performed simulation-based tests to examine the validity of our approach; we compare performance results obtained by simulation with the theoretical value obtained *a priori* from the scenario parameters and illustrate with a simple example a technique for estimating the required sampling rate to achieve a given level of performance.

Keywords: Hidden Markov Model, sampling, performance estimation, (α, β) -currency

1. INTRODUCTION

Many surveillance and tracking scenarios involve “observing” the real world at discrete intervals in time, obtaining (usually noisy) snapshots of certain parameters, and filtering these observations using a variety of algorithms to determine whether there exist certain processes that we are interested in, and if so, what their state is.

This abstract description is the basis for the concept of a Process Query System (PQS¹), a generic architecture for process tracking and sensor information fusion. Concrete examples of the types of processes that we may want to track include computer network attacks (such as internet worms) and battlefield scenarios (such as enemy tank and troop movements).

The sampling rate is the frequency at which observations are collected using our sensing system. This frequency may be fixed (observations are collected at constant intervals in time), or variable. The importance of the sampling rate lies in the fact that higher sampling rates allow for more accurate tracking, but also consume more resources. Therefore, our goal is to be able to specify a level of performance in some meaningful terms and then estimate what sampling frequency will be required to fulfill such performance requirement.

The process model that we have considered in this work is the Hidden Markov Model (HMM^{2,3}), because of its flexibility and mathematical structure, as well as its natural links to target tracking and information theory. Hidden Markov Models are currently employed in a wide variety of applications, including speech recognition³, target tracking⁴ and protein sequence analysis⁵. The Viterbi algorithm is perhaps the best known method for tracking the hidden states of a process from a sequence of observations.

There are several performance measures that seem suitable for process tracking. For example, the entropy of the observations provides a measure of the uncertainty in the observation sequence generated by a Hidden Markov Model. However, in a surveillance scenario there is another measure that is even more suitable to express

Further author information:

Diego Hernando: E-mail: diego.hernando@ieee.org

Valentino Crespi: E-mail: vcrespi@calstatela.edu

the performance of our system: the (α, β) -currency⁶⁻⁸ of a monitoring system provides a measure of how accurate and up-to-date the system's view of the world is. We extend the description of this concept later in this paper.

Another fundamental aspect associated with surveillance and tracking is the fact that, in many situations, there is not just one instance of a process that should be tracked, but rather there are several instances of different processes, evolving according to different models. Thus, in addition to estimating the evolution of each hidden process, we need to associate each received observation with its process¹¹. Data association adds another layer of difficulty to the task of the monitoring system, which should be taken into account when estimating performance and deriving sampling rate requirements.

The remainder of this document is organized as follows:

- Section 2 gives a brief overview of the background knowledge necessary to understand this work. This includes Hidden Markov Models, (α, β) -currency and Multiple Hypothesis Tracking.
- Section 3 describes Hidden Markov Models in more detail as they relate to the sampling frequency at which the observations are collected.
- Section 4 shows the connection between the concept of (α, β) -currency and the techniques for tracking processes using Hidden Markov Models.
- Section 5 describes the experiments that have been performed along with the theoretical work. In particular, we intend to prove that the measures of performance we derive actually apply in a few simple simulated scenarios.
- Section 6 summarizes the main results and lessons learned from this work.
- Section 7 explains the future lines of work that seem most promising, given the current state of our work.

2. PRELIMINARIES

2.1. Hidden Markov Models

A Hidden Markov Model (HMM) is, in the words of Ephraim,² a discrete-time finite-state homogeneous Markov chain observed through a discrete-time memoryless invariant channel. The channel is characterized by a finite set of transition densities indexed by the states of the Markov chain.

Formally, an HMM consists of the following elements:

- A Markov chain $\{S_i\}$, represented by an $N \times N$ stochastic matrix A , which describes the transition probabilities $a_{ij} = P(S_t=j|S_{t-1}=i)$ between the states of the model, together with a probability distribution π , where $\pi_i = P(S_1=i)$.
- A set of probability distributions, one for each hidden state, $b_i(o_j) = P(O_t=o_j|S_t=i)$, that model the emission of such observations. In this work we consider only finite observation sets of cardinality M , and so we can accommodate the probability distributions b_i in the rows of an $N \times M$ matrix B .

According to Rabiner³, there are three basic problems concerning Hidden Markov Models:

1. Given an observation sequence $O^t = O_1O_2 \cdots O_t$ and a model $\lambda = (A, B, \pi)$, computing $P(O^t|\lambda)$, i.e., the probability of the observation sequence given the model. This can be achieved efficiently using the *forward* (or *backward*) *algorithm*.
2. Given an observation sequence $O^t = O_1O_2 \cdots O_t$ and a model $\lambda = (A, B, \pi)$, finding a corresponding state sequence \hat{S}^t which is optimal (i.e. best "explains" the observations) in some meaningful way. One solution to this problem is the widely used *Viterbi algorithm*⁹.

- Adjusting the model parameters $\lambda = (A, B, \pi)$ for a given “training” observation sequence O^t , i.e., estimating $\text{argmax}_{\lambda}[P(O^t|\lambda)]$. There is no known efficient way to find the globally best parameters for an observation sequence. There are, however, iterative procedures to find a local maximum of $P(O^t|\lambda)$, such as the *Baum-Welch method*³.

Throughout this paper we will adopt the following notation:

- Subscripts will be used to identify a particular component in a sequence.
- Superscripts will be used to denote sequences of variables or symbols. For example, by S^t we will mean the sequence of t random variables (S_1, S_2, \dots, S_t) .
- Capital letters will be used to denote random variables while small letters will denote specific symbols of a probability source.

2.2. (α, β) -currency

The concept of (α, β) -currency⁷ was introduced as a measure of performance for monitoring systems. Let us assume there is a monitoring system that keeps track of the state of a collection of items. Each time the system “visits” (or samples) one of the items, it records its state. Thus, at any point in time, the monitor has a record of the state of each item, as it was when the item was last visited.

However, since the frequency with which the monitor samples the items is finite, an item may have changed its state since the last time it was visited, and therefore the information in the monitor is not “current” with respect to that item. This temporal aspect of what it means to be current can be relaxed by introducing a *grace period* β : an object’s index entry is β -current if the object has not changed between the last time it was re-indexed and β time units before the present. The smaller the value of β , the more current our information about the object is.

To determine whether an index entry for an object is β -current, Brewington⁶ defines the following:

- t_{Δ} is the most recent time at which the object’s state changed,
- t_n is the current time,
- t_0 is the last time the object was observed,
- T is the sampling period (revisit interval), therefore the object will next be observed at time $t_0 + T$,
- β is our grace period.

The probability that the index entry for an object is β -current at time $t_n > t_0 + \beta$ is

$$P(\text{a single object is } \beta\text{-current} | t_0, t_n) = 1 - P(t_0 \leq t_{\Delta} \leq t_n - \beta) \quad (1)$$

This probability is defined for a single object, but can be extended to the whole system (set of objects), by introducing the probability α of being β -current. We can assume that the objects in the system have a change rate λ , and are reindexed every T time units (both λ and T are random variables, generally with high correlation: the higher the change rate of an object, the smaller the sampling interval should be for that object). Let $h(\lambda, T)$ be the joint probability density for (λ, T) . This is assumed to be independent of t_n , which is distributed according to a density $x(t_n)$. Finally, define the probability $P(\text{a single object is } \beta\text{-current} | \lambda, T, t_n)$ of a single entry being β -current given λ, T, β and the time t_n . Thus, the probability α that the system is β -current is:

$$\begin{aligned} \alpha &= P(\text{the index is } \beta\text{-current}) \\ &= \int_T \int_{\lambda} \int_{t_n} [P(\text{a single object is } \beta\text{-current} | \lambda, T, t_n) x(t_n) dt_n] h(\lambda, T) d\lambda dT, \end{aligned} \quad (2)$$

and the monitoring system is said to have (α, β) -currency.

2.3. Multiple Hypothesis Tracking

The problem of tracking a single process from a sequence of observations (a problem which can sometimes be stated using Hidden Markov Models or, in the case of physical targets with Gaussian noise in a continuous state space, with Kalman Filters¹⁰), can be extended to the multiple target case by adding one dimension of complexity, due to having to find the best association between observations and targets.

In 1979, Reid¹¹ proposed his Multiple Hypothesis Tracking (MHT) algorithm for tracking multiple targets and handling the creation of tracks, the assignment of observations to existing tracks, and the discarding of observations as noise from the sensing system. In this context, a track is a sequence of observations estimated to have been produced by the same process. A hypothesis consists of a set of consistent (non-contradictory) tracks obtained at some point during the tracking process. Obviously the total number of hypotheses that explain the observations will grow without bound as more observations are made by the system. For example, assuming a reduced system in which the only possibilities are to consider each observation as a new target or as clutter, the number of hypotheses after n observations will be 2^n . To avoid this exponential growth, a process called suboptimization (or pruning) is applied after each observation (or observation vector) is received in order to keep only the k most significant hypotheses.

The filter in Reid’s MHT generates *measurement-oriented* hypotheses to account for all possible origins of a measurement. This means that for each observation that is given to the algorithm, every possible target (existing track) is examined for this measurement, and also the possibilities of the observation corresponding to a new target or clutter are examined.

Reid derived the algorithm’s equations in terms of Kalman Filters, but they can be restated using other process models. In this work, we apply MHT using the probabilities obtained from Hidden Markov Models instead of the likelihoods obtained from Kalman Filters.

3. HIDDEN MARKOV MODELS: CATEGORIES AND PERFORMANCE

In relation to the sampling rate, we can distinguish between two types of HMMs, depending on the kind of real-world process they model. We consider the following two cases:

- The state transitions occur at fixed intervals in time. For example, we can have a communications system in which the symbols to be transmitted follow a Markov chain X_n in the space \mathcal{X} of hidden states, and a new symbol is produced every second. The symbols are sent through a noisy channel such that the probability distribution of the output Y_n , output symbol in the space \mathcal{Y} (consisting, in the discrete case, of M distinct observation symbols), depends only on the current input X_n . This process can be naturally represented with a Hidden Markov Model where the input symbols correspond to the hidden states in the model, the transitions between them are modeled by the transition matrix A , and the emission of observations is modeled by the emission probability matrix B , which depends on the channel. We may raise the sampling rate in this process by transmitting the input symbols more than once per second. We shall focus in the cases where these symbols are independent and identically distributed. By doing so, we would modify our model only in its emission probabilities, i.e., we would obtain several output symbols per input symbol. These could be combined into an output vector of symbols. If we obtain k observations per second, this output vector lies in the space \mathcal{Y}^k , and we obtain the probability distribution over \mathcal{Y}^k by forming an augmented emission matrix B^{aug} where each row is the result of k Kronecker products of the original row in B with itself. If $b_i = \{b_{i,1}b_{i,2}\dots b_{i,M}\}$ is a row in the emission matrix B ,

$$b_i^{aug} = b_i \otimes b_i \otimes \dots \otimes b_i . \tag{3}$$

Intuitively, such an emission matrix would amplify the differences between the observation emission probability distributions associated with the different input symbols. We can consider the relative entropy¹² between the observation emission probabilities between any two distinct input symbols, represented in rows i and j of the original emission probability matrix B . The relative entropy between these two probability distributions is $K(b_i, b_j) = \sum_{r=1}^M b_i(r) \cdot \log\left(\frac{b_i(r)}{b_j(r)}\right)$. The distance between the probability distributions if the sample rate is k times higher would be:

1. *Fixed state transition time.* In this category we place HMMs whose Markov chains change state at regular intervals of time Δt . In those cases, the probability that the state of the system does not change in the interval $[t_0, t_n - \beta)$ will be

$$\forall u, t_0 \leq u \leq t_n - \beta : P(S_u = S_{t_0} | O^{t_0}) = \sum_{i=1}^N (a_{i,i})^k \cdot P(S_{t_0} = i | O^{t_0}), \quad (6)$$

where

$$k = \begin{cases} \left\lfloor \frac{t_n - \beta - t_0}{\Delta t} \right\rfloor, & t_n \geq t_0 + \beta \\ 0, & \text{otherwise} \end{cases}$$

2. *Stochastic state transition time.* In this category³ we place HMMs in which the time between two state transitions is a random variable. In this case the above probability becomes:

$$\forall u, t_0 \leq u \leq t_n - \beta : P(S_u = S_{t_0} | O^{t_0}) = \sum_{i=1}^N a_{i,i}(t_n - \beta - t_0) \cdot P(S_{t_0} = i | O^{t_0}). \quad (7)$$

Therefore, the (α, β) -currency is determined by the parameters of the HMM, plus the probability distribution $P(S_K = i | O^K)$, $1 \leq i \leq N$. However, the expected value (in the limit as $K \rightarrow \infty$) of this probability distribution is computable *a priori*, and therefore so is the expected value of α given β .

$$\begin{aligned} E_K [P(S_K = i | O^K)] &= \lim_{t \rightarrow \infty} \left[\frac{1}{t} \sum_{k=1}^t P(S_k = i | O^k) \right] \\ &= \lim_{t \rightarrow \infty} P(S_t = i | O^t) \\ &= \lim_{t \rightarrow \infty} \sum_{o^t \in \mathcal{O}^t} [P(O^t = o^t) \cdot P(S_t = i | O^t = o^t)] \\ &= \lim_{t \rightarrow \infty} \sum_{o^t \in \mathcal{O}^t} P(S_t = i, O^t = o^t) \\ &= \lim_{t \rightarrow \infty} P(S_t = i), \end{aligned} \quad (8)$$

where the second step is due to the *Cesàro mean*. The resulting expression is equal to p_i , the stationary distribution of the hidden Markov chain. This distribution exists if the model is ergodic², and it is computable *a priori* simply from A , the transition matrix of the hidden states. Therefore, the expected value of α , given β , the model λ , and the sampling rate f_s ($E[\alpha | \beta, \lambda, f_s]$), can be computed from the stationary distribution of the Markov chain.

Being able to predict the (α, β) -currency *a priori* allows us to determine which sampling rate will be necessary to obtain a certain level of performance in monitoring the process. For a certain performance requirement (α_0, β_0) , we can start with a random sampling frequency $f_{s,1}$ and compute $E[\alpha_1 | \beta_0, \lambda, f_{s,1}]$. If the value obtained is smaller than α , we choose a higher sampling rate and repeat the process until we reach $\alpha_i \geq \alpha_0$, then choose that last frequency $f_{s,i}$. If $\alpha_1 > \alpha_0$, we reduce the frequency and repeat the process until $\alpha_i < \alpha_0$, and then choose the previous value $f_{s,i-1}$.

4.1.2. Second approach: the state remains as estimated

This second notion of (α, β) -currency imposes a stronger condition. In this case, α represents the probability that the state remains as estimated the last time we sampled (with a grace period of β). Therefore, we not only require that the hidden state has not changed, but we also include the probability that our estimation was correct.

$$\alpha = P[(\hat{S}_{t_0} \text{ was correct}) \cap (\text{state is maintained up to } t_n - \beta)], \quad (9)$$

where \hat{S}_{t_0} is the state estimation at time t_0 , $\hat{S}_{t_0} = \operatorname{argmax}_i P(S_{t_0}=i|O^{t_0})$.

This approach for determining the tracking performance seems especially useful in situations where there are certain hidden states that require an alarm to be raised. Using this technique, we could refine the measure of α by estimating the probability that we are in one of these troublesome hidden states at any time after we last sampled the system. During normal operation, we could optimize our system resources by not sampling at a constant rate, but rather computing the (α, β) -currency at runtime and querying our sensing system for a new sample as soon as α reaches a certain threshold value. In other words, if we have some states considered *safe*, plus some *unsafe* states, we could redefine α as the probability of entering an unsafe state up to β time units ago.

In this case, the condition for models where the state transitions occur at regular intervals in time is defined as:

$$\forall u, t_0 \leq u \leq t_n - \beta : P(S_u = \hat{S}_{t_0} | O^{t_0}) = (a_{\hat{S}_{t_0}, \hat{S}_{t_0}})^k \cdot P(S_{t-f-1} = \hat{S}_{t_0} | O^{t_0}),$$

where \hat{S}_{t_0} is the index of our state estimation (most likely state according to the model) at the previous sampling time, given all the observations received up to that time:

$$\hat{S}_{t_0}(O^{t_0}) = \operatorname{argmax}_i P(S_{t_0}=i|O^{t_0})$$

and

$$k = \begin{cases} \left\lfloor \frac{t_n - \beta - t_0}{\Delta t} \right\rfloor, & t_n \geq t_0 + \beta \\ 0, & \text{otherwise} \end{cases}$$

In this case, the condition on the optimal sampling frequency can be expressed modifying (5) as follows (note the inclusion of \hat{S}_{t_0} instead of S_{t_0} inside the expectation, specifying the added condition that our estimate was correct):

$$f_{\mathcal{M}}(\beta) = \min_f \{ E_{t_n} [P(S_u = \hat{S}_{t_0} | O^{t_0}), \forall u : t_0 \leq u \leq t_n - \beta] \geq \alpha, t_0 \leq t_n \leq t_0 + f^{-1} \}. \quad (10)$$

4.2. (α, β) -currency when tracking multiple processes

The notion of (α, β) -currency extends naturally to sets of independent HMMs. Formally, let Θ be a hypothesis of n consistent and independent tracks each modeled by an HMM \mathcal{M}_i . Then we say that Θ is β -current if all the HMMs associated with its tracks are β -current. Furthermore, if each HMM \mathcal{M}_i is (α_i, β) -current then, under the assumption of stochastic independence, we can say that Θ is $(\prod_i \alpha_i, \beta)$ -current.

5. SIMULATIONS

5.1. First notion of (α, β) -currency

We have performed simulations to compute the (α, β) -currency of simple HMMs, as defined in section 4.1.1, for different values of β and different sampling rates. We compare our results to the predicted values for the (α, β) -currency, as described in section 4.1.

The values of α were obtained averaging (1) over one sampling interval:

$$P(\text{a single object is } \beta\text{-current} | t_0) = \frac{1}{T_s} \int_{t_0}^{t_0+T_s} P(\text{a single object is } \beta\text{-current} | t_0, t_n) dt_n, \quad (11)$$

where $P(\text{a single object is } \beta\text{-current} | t_0, t_n)$ can be computed using (6) if the state transitions occur at constant intervals, or (7) if the transitions occur at random times. It must be noted that when we compute the integral in this expression, the probability that a single object is β -current at a time $t_n < t_0 + \beta$ is 1, as explained in section 4.1.

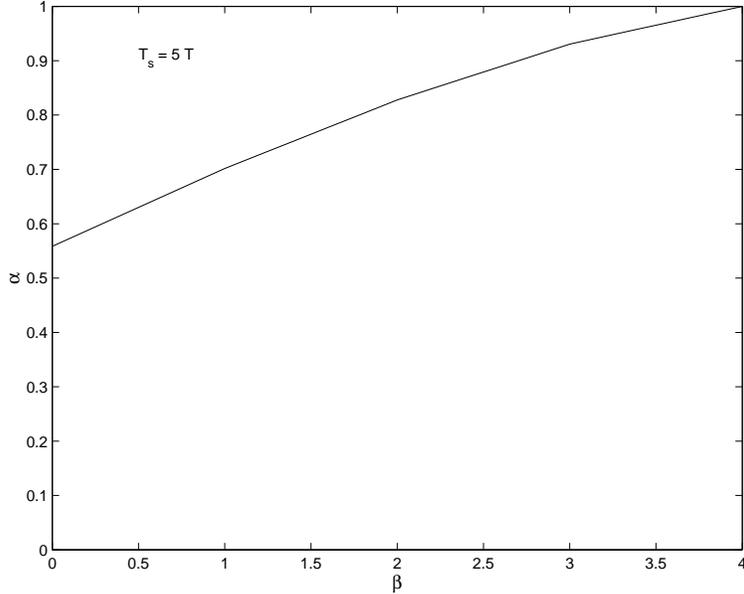


Figure 1. Relation between α and β , for fixed sampling interval. The sampling interval is $T_s = 5 \cdot T$, where T is the interval at which the state jumps are produced in the model.

Fig. 1 shows the relation between α and β for a particular sampling rate. The average values of α obtained experimentally converge to their predicted “stationary” values in the limit as the length of the observation sequence grows (see section 4.1). Furthermore, the differences that we have recorded between predicted and experimental values of α become negligible after only a few observations.

Fig. 2 shows the relation between α and the sampling interval $T_s = f_s^{-1}$, for a given value of β . This is an example of how the sampling rate can be adjusted to achieve a given level of tracking performance. Let us assume that the performance we wish to achieve for a certain application is (0.8, 1.0)-currency, i.e., we require to be 80% sure at any time that the hidden state has not changed since our last estimation. Then, according to Fig. 2, it would suffice to sample with an interval $T_s \simeq 3.8 \cdot T$.

5.2. Alternative notion of (α, β) -currency

Fig. 3 shows the (α, β) -currency following the approach described in section 4.1.2. This second approach is more restrictive than the first, since it not only imposes that the state does not change between the time when the target was last observed (t_0) and the current time minus the grace period ($t_n - \beta$), but also that our state estimate at time t_0 was right. The α obtained with the first approach could always be made as close to 1 as desired by sampling at a higher rate or relaxing the grace period. However, in this case, even if $t_n = t_0$ the value of α will only be as high as $\max_i P(S_{t_0} | O^{T_0})$, i.e., the probability that our estimation is correct. Therefore, if we cannot achieve the desired (α, β) -currency simply by examining the sampling rate, we will have to go back to our model and, for example, implement a better (less noisy) sensing system that makes the emission matrix B less ambiguous.

In the cases where this is possible, we can also adopt one of the techniques described in section 3: sampling at a rate higher than the rate at which the state jumps occur, and therefore obtain more than one observation per state. Whether or not this is possible depends on the nature of the process being modeled. If it is possible, then it’s effectively like increasing the granularity of the observation space and increasing the distance between the probability distributions of the observations generated by the different states. This would be conceptually equivalent to improving the sensing system to produce better observations.

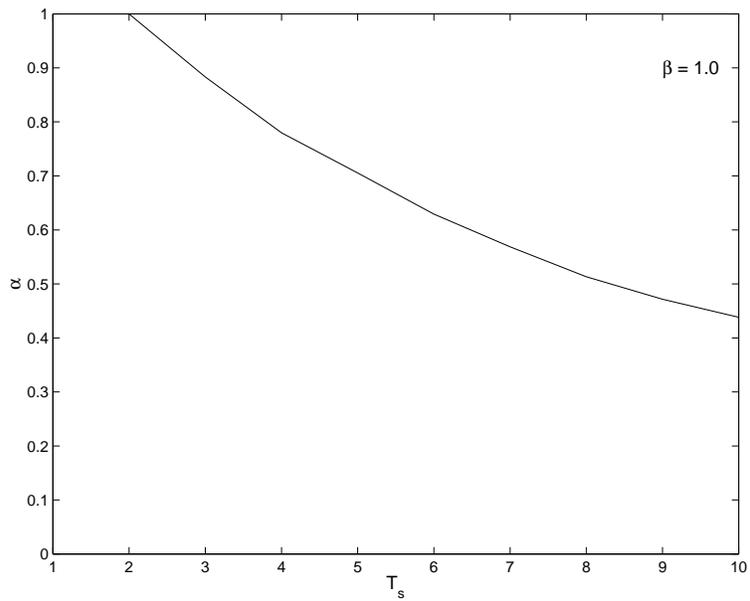


Figure 2. Relation between α and T_s , for fixed $\beta = 1.0$ and $T = 1.0$.

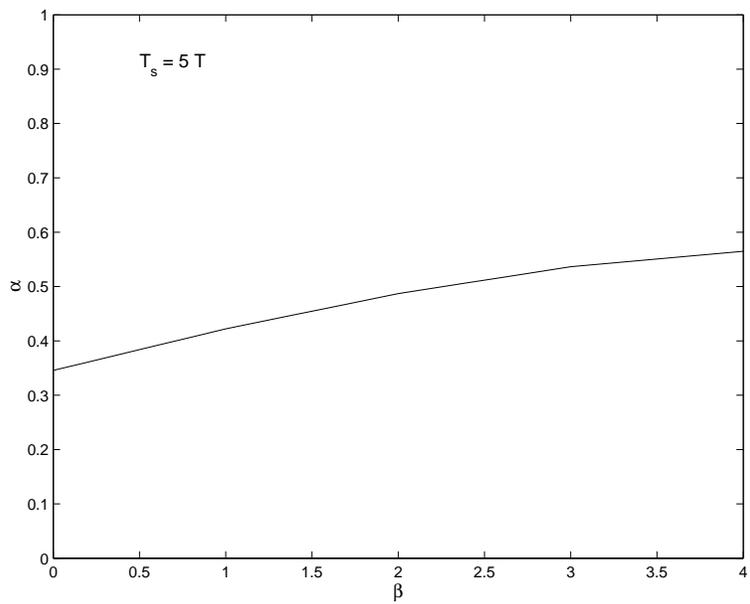


Figure 3. Relation between α and β when applying the second approach for determining (α, β) -currency

6. CONCLUSIONS

We have applied a measure of performance for tracking processes with HMMs. This allows us to relate the quality of the results that we wish to obtain to the sampling rate at which observations are collected.

In some cases, this allows us to predict which sampling frequency will be required to achieve a given performance level. In general, this measure can be used to implement a monitoring system that dynamically decides when it is time to query a new set of observations from the sensing system.

The single target case is rather straightforward to conceive. The same concepts extend naturally to the more interesting multitarget scenario, although the design of measures and algorithms are somewhat more complex, due to the extra layer of complexity related to data association.

7. FUTURE WORK

We are currently working to further develop the multitarget case. The concept of (α, β) -currency should allow us, also in this case, to assure a certain tracking performance while minimizing the use of resources (where by resources we understand a set of different factors like network usage, power consumption in the sensors, computational complexity in the algorithm, ...)

Moreover, we are searching for better ways to characterize *a priori* the (α, β) -currency of an arbitrary HMM, following the *second* approach (section 4.1.2). This would imply computing (perhaps by simulation) *a priori* the expected quality of the estimates produced with a model.

ACKNOWLEDGMENTS

The authors wish to thank George Cybenko for his guidance during this work and for the fruitful discussions that resulted in most of the ideas presented in this paper. The authors would also like to thank Guofei Jiang, Alex Jordan and Yong Sheng for their reviews and comments on the manuscript.

This research was partially supported by: DARPA project F30602-00-2-0585; National Institute of Justice, Department of Justice award number 2000-DT-CX-K001.

REFERENCES

1. V. Berk, W. Chung, *et al.*, "Process Query Systems for Surveillance and Awareness," in *7th World Conference on Systemics, Cybernetics and Informatics (SCI2003)*, July 2003.
2. Y. Ephraim and N. Merhav, "Hidden Markov Processes," *IEEE Transactions on Information Theory* **48**, June 2002.
3. L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE* **77**(2), 1989.
4. H. A. P. Blom and Y. Bar-Shalom, "The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients," *IEEE Transactions On Automatic Control* **33**(8), 1988.
5. A. Krogh, M. Brown, I. S. Mian, K. Sjolander, and D. Haussler, "Hidden Markov Models in Computational Biology: Applications to Protein Modeling," Tech. Rep. UCSC-CRL-93-32, University of California at Santa Cruz, 1993.
6. B. Brewington, *Observation of changing information sources*. PhD thesis, Dartmouth College, June 2000.
7. B. E. Brewington and G. Cybenko, "Keeping up with the changing Web," *Computer* **33**(5), pp. 52–58, 2000.
8. B. E. Brewington and G. Cybenko, "How dynamic is the Web?," *Computer Networks (Amsterdam, Netherlands: 1999)* **33**(1–6), pp. 257–276, 2000.
9. G. D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE* **61**(3), 1973.
10. R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, third ed., 1997.
11. D. B. Reid, "An Algorithm for Tracking Multiple Targets," *IEEE Transactions on Automatic Control* **24**(6), 1979.
12. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, John Wiley & Sons, New York, NY, USA, 1991.