The Forward Procedure for HSMMs based on Expected Duration

Jan Lemeire, Francesco Cartella

***Abstract*—For dynamic models, the forward procedure is used to calculate the probability of an observation sequence for a given model. For Hidden Semi-Markov Models (HSMMs), the calculation can be approximated by keeping track of the expected state duration instead of the distribution. The update equation for the expected duration proposed by Azimi et al.[1] is, however, wrong. The experiments presented by Azimi et al.[1] did not reveal the error, since for the presented cases the state duration does not play a role in the probabilities. We propose a better equation for updating the expected duration. It nevertheless remains an approximation for calculating the probability of observation sequences. We analyze the assumptions to show under which conditions the approximation errors become important. Experiments show that the approximation is only reasonable for left-2-right HSMMs. As we focus on a specific sub class of HSMMs we derive specialized equations from the general form for the exact calculation of the forward variable.**

***Index Terms*—HSMMs, semi-Markov models, forward procedure.**

$\mathcal{S}$	$=\{1,2,\ldots,Q\}$, the state space
$\mathcal{D}$	$=\{1,2,\ldots,D\}$, all possible state durations
$\mathcal{V}$	$=\{v_1,v_2,\ldots,v_K\}$, the observation space
$s_{1:T}$	$\triangleq s_1,s_2,\ldots,s_T$, the state sequence, $S_t \in \mathcal{S}$
$o_{1:T}$	$\triangleq o_1,o_2,\ldots,o_T$, the observation sequence, $o_t \in \mathcal{V}$
$s_{[t_1:t_2]}=i$	state i started at time t_1 and lasts till t_2 with duration $d=t_2-t_1$. It implies $s_{t_1-1}\neq i$ and $s_{t_2+1}\neq i$.
$s_{[t_1:t_2}=i$	state i started at time t_1 and lasts till t_2. It implies $s_{t_1-1}\neq i$ while s_{t_2+1} may or may not be i.
$s_{[t_1}=i$	state i started at time t_1 : $s_{t_1-1}\neq i$ and $s_{t_1}=i$
$d_t(i)$	time spent in state i prior to time t
$b_i(v_k)$	observation probabilities
$\phi_i(d)$	duration distribution
$\Phi_i(d)$	cumulative duration distribution, derived from $\phi_i(d)$
$r_i(d)$	recurrent state change probabilities, derived from $\phi_i(d)$
$a^0_{i,j}$	non-recurrent state transition probabilities

TABLE I
HSMMS: NOTATIONS AND DEFINITIONS

I. INTRODUCTION

HIDDEN Semi-Markov Models (HSMMs) provide a useful extension to Hidden Markov Models (HMMs) by taking into account the time the system resides in a certain state (the sojourn time or duration) for determining whether a state transition will take place. The main assumption on the system changes from 'future is independent from past given the current state' to 'future is independent from past given the current state and the state duration'. The latter defines HSMMs in their most general form [2]. We focus on a widely-used HSMM subclass [1], [3] which employs 3 assumptions:

1) state duration is independent from the previous state,
2) given a state change, the probability of entering a certain state is independent from the duration of the previous state,
3) the observation probability only depends on the current state.

Usage of HSMMs can be done according to the same principles as developed by Rabiner [4] for HMMs. The probability of an observation sequence given a model ($P(O|\lambda)$) is calculated by induction [4]. One defines a forward variable which is updated for each additional time step. For HMMs, we have to keep track of the state beliefs ($\alpha_t(i) \triangleq P(s_t=i,o_{1:t}|\lambda)$). For HSMMs, we have to retain the probability of being in a state for a certain period [2, Section 2.2.1]:

$$\alpha_t(i,d) \triangleq P(s_{[t-d+1,t]}=i,o_{1:t}|\lambda) \tag{1}$$

The authors are with the Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium. Jan Lemeire is also with the Dept. of Industrial Sciences (INDI) and with the Dept. of Multimedia Technologies, iMinds, Technologiepark 19, 9052 Zwijnaarde, Belgium (e-mail: jan.lemeire@vub.ac.be).

The notations are defined in the next section.

As this procedure becomes computationally intensive, Azimi et al.[1] propose to keep track of $\alpha_t(i)$ (as with HMMs) and the expected duration of a state $\hat{d}_t(i)$ instead of $\alpha_t(i,d)$. The forward procedure is then based on iteratively updating $\alpha_t(i)$ and $\hat{d}_t(i)$. As we will discuss, the computational complexity decreases to a quadratic performance. However, the equation to calculate $\hat{d}_{t+1}(i)$ from Azimi et al.[1] makes no sense. With a simple example it can be shown that it produces large errors on the state probabilities $\alpha_t(i)$. Exactly in those cases where the state cannot be derived from the observations and hence state duration gives information on state changes, Azimi's approximative equations appeared to be wrong. We propose an alternative equation, which is also an approximation, but which is the best one can do given the information. For a correct calculation of $\hat{d}_t(i)$, one would need to use Eq. 1 and the exact forward procedure.

In the next section we give the definition of the HSMM sub class we focus on. We derive the exact forward procedure and the approximate procedure based on expected duration with our alternative update equation. Next, the failure of Azimi's equation is discussed. Then we analyze the errors made by the approximate procedure. Before concluding we compare the runtime complexities of each approach.

II. HSMM SUBCLASS DEFINITION

We adopt the same notations and definitions as Yu [2]. They are listed in Table I.

According to Yu's general definition, a state and its duration depend on the previous state and its duration (first equation of Section 2.1). In our case, state duration is independent from

the previous state (Assumption 1). Then Eq. 1 of Yu [2] can be applied (in the sense that this is what we need to know about the state change probabilities):

$$a_{(i,d)j} \triangleq P(s_{[t+1} = j | s_{[t-d+1:t]} = i) \tag{2}$$

In contrary to Yu [2] we also define a for $i = j$ since we will turn the HSMM into an HMM with dynamic transition probabilities which are calculated from the expected duration. Another simplifying assumption often employed [1], [3] is that given a state change, the probability of entering a certain state is independent from the duration of the previous state (Assumption 2). As a result, the transition probabilities can be defined by a **duration distribution** $\phi_i(d)$ and a **base transition matrix** $a^o_{i,j}$ containing the non-recurrent state transition probabilities. The state transition probabilities $a_{(i,d)j}$ are then defined as follows (which is equivalent to Eq. 2 of [1]):

$$a_{(i,d)j} = \begin{cases} r_i(d) & \text{if } i = j \\ (1 - r_i(d)).a^0_{ij} & \text{if } i \neq j \end{cases} \tag{3}$$

with $r_i(d)$ the recurrent transition probabilities (the probability that the system remains in the same state). They are defined and derived from the duration distributions as follows [1, Eq. 2]:

$$r_i(d) \triangleq P(s_t = i | s_{[t-d:t-1} = i) \tag{4}$$

$$= (1 - \Phi_i(d))/(1 - \Phi_i(d-1)) \tag{5}$$

where Φ_i denotes the cumulative density function of state i's duration distribution ϕ_i.

Note that this was incorrectly defined as $1-\Phi_i(d-1)$ in [5, Eq. 6] but corrected later [1]. The derivation of Eq. 5 follows from the definition of the duration distribution.

$$r_i(d) \triangleq P(s_t = i | s_{[t-d:t-1} = i) \tag{6}$$

$$= P(s_{[t-d:t} = i | s_{[t-d} = i)/P(s_{[t-d:t-1} = i | s_{[t-d} = i) \tag{7}$$

Nominator and denominator can be expressed as a function of Φ:

$$\phi_i(d) \triangleq P(s_{[t-d:t-1]} = i | s_{[t-d} = i) \tag{8}$$

$$\Phi_i(d) = \sum_{d'=1}^{d} P(s_{[t-d:t-d']} = i | s_{[t-d} = i) \tag{9}$$

$$\Rightarrow P(s_{[t-d:t-1} = i | s_{[t-d} = i) = 1 - \Phi_i(d-1) \tag{10}$$

III. The exact forward procedure

We first derive from the general equations the equations for the forward procedure under the 3 assumptions. We will employ a forward variable which is a bit different than Eq. 11 used by [2]:

$$\alpha_t(i,d) \triangleq P(s_{[t-d+1:t} = i, o_{1:t} | \lambda) \tag{11}$$

For general HSMMs, one should employ $P(s_{[t-d+1:t} = i], o_{1:t} | \lambda)$. However, since the duration is independent from the previous state and duration (Assumption 1), we can include the states that will remain in the state after time t (and ommit the right bracket).

We first explicitly calculate $\nu_t(j,d) \triangleq P(s_{[t-d+1:t} = j, o_{1:t-1})$ in the following way:

$$\nu_t(j,d) = \begin{cases} \sum_{i \in \mathcal{S} \setminus \{j\}} \sum_{d' \in \mathcal{D}} a_{(i,d')j} \alpha_{t-1}(i,d') & \text{if } d = 1 \\ a_{(j,d-1)j} \alpha_t(j, d-1) & \text{if } d > 1 \end{cases} \tag{12}$$

then:

$$\alpha_t(j,d) = \nu_t(j,d) \cdot P(o_t | s_t = j) \tag{13}$$

where the last factor is the observation likelihood b_i, which only depends on the current state (Assumption 3).

IV. The forward procedure based on expected duration

Azimi et al. [1] proposed a faster and less cumbersome forward procedure based on expected duration of a state:

$$\hat{d}_t(i) \triangleq \mathbb{E}(d_t(i) | s_t = i, O_{1:t}) \tag{14}$$

The expected duration is the expected time the system was in state i given that it is currently in state i. Note that Yu [2] defines $\hat{d}$ as the Maximum A Posteriori (MAP) of the duration (Eq. 10) for being used in the EM algorithm. The state transition matrix is then calculated based on the expected state durations:

$$a_{(i,d)j} \approx a_{(i,\hat{d})j} = \begin{cases} r_i(\hat{d}(i)) & \text{if } i = j \\ (1 - r_i(\hat{d}(i))).a^0_{ij} & \text{if } i \neq j \end{cases} \tag{15}$$

The forward procedure will now keep track of the expected duration and the forward variable $\alpha_t(i,d)$ is replaced with $\alpha_t(i) \triangleq P(s_t = i, o_{1:t} | \lambda)$ (which is the same as the one used for HMMs). In practice we have to overcome underflows since the forward variable exponentially goes to zero. Therefore we employ the following forward variable [6], [2], [7]:

$$\bar{\alpha}_t(i) \triangleq P(s_t = i | o_{1:t}, \lambda). \tag{16}$$

This variable can easily be calculated by normalizing the original α_t after each iteration.

$P(s_t = i, o_t | o_{1:t-1})$ is calculated and normalized by dividing the probabilities by factor $P(o_t | o_{1:t-1})$. Note that this factor is needed in the calculation of $\hat{d}_{t+1}$ (see Appendix B).

We propose the following equation to update the average duration (although it is also an approximation):

$$\hat{d}_{t+1}(i) \approx P(s_t = i | s_{t+1} = i, o_{1:t+1})(\hat{d}_t(i) + 1) \tag{17}$$

The intuition behind the equation is that the current average duration is the previous average duration plus 1 weighted with the 'amount' of the current state that was already in state i in the previous step. The derivation, the approximations and the equations for practical usage are given in the appendices.

V. The error of Azimi's update equation

Azimi et al. proposed the following update equation [1, Eq. 7]:

$$\hat{d}_{t+1}(i) = 1 + P(s_t = i | \theta, o_{1:t}) \cdot \hat{d}_t(i) \tag{18}$$

This equation is correct when $P(s_t = i | o_{1:t})$ is 1 or 0, since the duration is then incremented by 1 in the former or remains

state	$i=1$	$i=2$	$i=3$
observation distr $b_i(v_k)$	Gauss(10, 10)	Gauss(20, 10)	Gauss(22, 10)
duration distr $\phi_i(d)$	Gauss(25, 5)	Gauss(25, 5)	Gauss(25, 5)

TABLE II
PARAMETERS OF EXAMPLE LEFT-TO-RIGHT HSMM. GAUSS(μ, σ) REPRESENTS A GAUSSIAN DISTRIBUTION WITH MEAN μ AND VARIANCE σ.

1 in the latter. On the other hand, when the state is unsure, the duration is incremented with a value smaller than 1. However, the state uncertainty does not play a role in the expected duration, since the duration is the time the system is in a certain state, *would the system still be in state i at time t*. By the conditional definition it is independent from the state probability at time t. Assume for instance that the probability is zero that the system changes from a state $j \neq i$ into state i. Then, if the system is in state i at time t, it was already in state i at time $t-1$, so the duration should be incremented with 1.

That the error occurs when the current state is not exactly known from the observations is shown in Fig. 1. It shows a simulation of a 3-state left-to-right HSMM with the parameters of Table II. In a left-to-right HSMM, state i deterministically changes into state $i+1$. The final state is permanent. The state evolution and observations are shown. Additionally, the results of the forward procedure are shown (estimation of state based on the observations): the state probabilities $\bar{\alpha}$ and the expected duration $\hat{d}$. The state change from 1 to 2 is captured correctly by [1]'s equations, while the change to 3 is not. The first happens correctly because from the clear difference in observations, it ensures there is almost no uncertainty on the state. This is not true anymore when changing from 2 to 3: the state is quite uncertain because the observation probabilities of both states are very similar. We see in the lower plot that the expected duration based on [1] starts to decrease due to this state uncertainty. Since the probability mass of entering 2 lies almost entirely around time 25, this is clearly wrong: the expected state duration can only increase after the transition period. This error did not become apparent in the experiments of [1], since the observation distributions were quite different for each state (the distances between the mean of the Gaussians was 10 while the variance was also 10).

VI. APPROXIMATION ERRORS

We start with a small example to show the errors based on the approximation of Eq. 17. Consider a 3-state left-to-right HSMM with duration of 2 or 3 (both have a probability of 0.5) for states 1 and 2. Consider the same observation distribution for the 3 states, such that the observations do not provide information on the state probability. $P(s_t)$ is shown in Fig. 2. The thickness of the gray rectangle indicates the probability. The expected state duration is also shown. An important point is that if the state flipped from 1 to 2 at time 2, then it will flip to 3 at time 4 or 5. However, at time 4, the expected duration is 1.5. Based on this expected duration (Eq. 15) no state change will happen (only when the expected duration is 2 or 3).

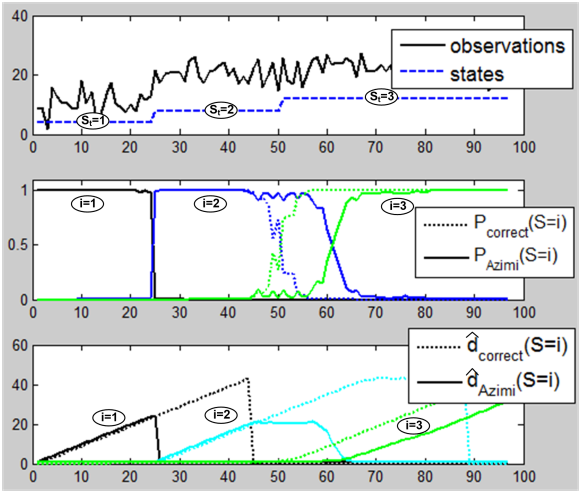


Fig. 1. Evolution of a 3-state left-to-right HSMM: observations and states. Based on the observations only, the true values (dotted line) and the values according to Azimi et al.[1] (solid line) for the state probabilities $\bar{\alpha}_t(i)$ and the expected duration $\hat{d}_t(i)$. The annotations show to which state the curve belongs.

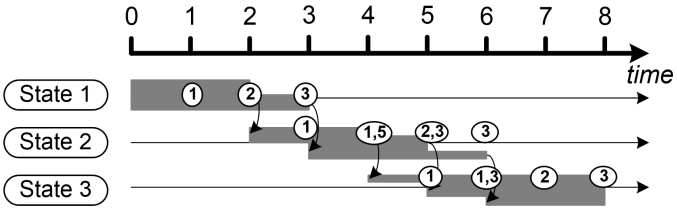


Fig. 2. State probabilities for a 3-state left-to-right HSMM with duration of 2 or 3 (both have a probability of 0.5) for states 1 and 2. The thickness of the gray rectangle indicates the state probability. expected state duration is indicated within the circles.

Our update equation is based on 2 approximations: Eq. 25 and Eq. 36 of the Appendix:

$$\begin{aligned} &P(s_{[t+1-d:t-1} = i | s_t = s_{t+1} = i, o_{1:t+1}) \\ &\approx P(s_{[t+1-d:t-1} = i | s_t = i, o_{1:t}) \end{aligned} \tag{19}$$

$$\begin{aligned} P(s_{t+1} = i | s_t = i, o_{1:t}) &= \sum_{d_t} r_i(d_t(i)).P(d_t | o_{1:t}) \\ &\approx r_i(\hat{d}_t(i)) \end{aligned} \tag{20}$$

It turns out that they have an opposite effect: the first results in an overestimation of the non-recurrent state transition probabilities, while the second results in an underestimation. Both effects cancel each other out after the transition period, as is shown in Fig. 3. It is a left-to-right HSMM with the same observation distribution for the first three states. The expected duration of states 1 and 2 is 100 and the variance of the duration is 35 and 5 respectively.

The forward procedure is correct during the transition from 1 to 2, since there is no uncertainty on the duration of state 1. The approximations start to deviate from the true values in the transition period from state 2 to 3. In the beginning, the second approximation underestimates the state change probability since the transition probability of the expected duration is different from the transition probability calculated

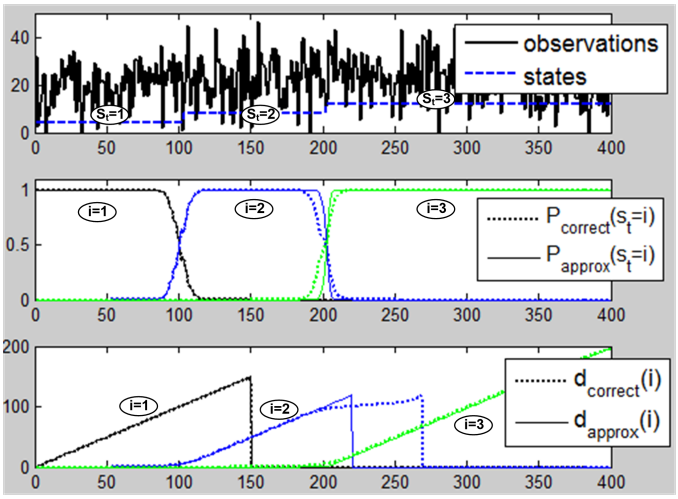


Fig. 3. Approximation error of left-to-right HSMM. The state probability $\bar{\alpha}_t(i)$ and the expected duration $\hat{d}_t(i)$ is plotted for Yu's equations and for the approximation based on expected duration.

as the sum over the products of state sojourn time distribution and duration distribution (Eq. 36).

This is also apparent in Fig. 2: when a state was entered early, the chance of leaving the state is greater than expected.

One clearly see that the true expected duration increases only sub-linearly when the chance of flipping has become non-zero. State 3 probability increases slower for the expected duration approximation because of the first error: the low variance of 3 applied on the expected duration gives a different transition probability than calculating it as the integral because for the latter, the flipping of the 'old durations' is taken into account.

The first approximation has the opposite effect. In those cases were the state is already for a long time in state i (long durations), the probability that it leaves the state is higher than expected. Since we update the expected duration with the previous expected duration, we do not take this early-entrance-early-leave effect into account. As a result, the expected duration is overestimated and therefore also the state change probabilities. This happens more toward the end of the transition region. In Eq. 24, we consider $P(s_t = i|s_{t+1} = i, o_{1:t+1})$ separately and then ignore $s_{t+1} = i$ from the second factor. As such, we are considering the transitions from $s_t = i$ to $s_{t+1} = i$ without taking the duration of state i into account.

We observed empirically that after the transition region, the approximation converges to the true value. Both approximations seem to cancel each other out. Because it is hard to prove in an analytical way that the error is bounded, we will rely on experiments. Table III shows the results. The approximation error of $\bar{\alpha}_t(i)$ and $\hat{d}_t(i)$ is measured during the run of a randomly-generated HSMM. The table shows the ranges from which the HSMM parameters are randomly chosen. Then, based on a random sample of 500 time steps generated from the HSMM, the average difference is measured between the true and approximated state distribution and expected duration of all states at all time steps. We also report the maximal difference. The error on the expected duration was only calculated when the state probability was larger than 0.001. The error can become very large when the state probability is low, but in those cases the expected state duration is irrelevant. The table shows the average over 100 experiments. We compare general HSMMs with left-2-right HSMMs. As was expected, the approximation errors become more significant when the observation distributions of the different states are very similar. For left-2-right HSMMs the errors remain within bounds due to the cancellation effect. In general, however, the approximation cannot be trusted.

Forward procedure	complexity	order
Yu's general equations	$Q.Q.D.D$	$O(D^2.Q^2)$
Equations with the 3 assumptions	$Q.(D+Q.D)$	$O(D.Q^2)$
Our update equation for $\hat{d}$	$Q.Q+6Q$	$O(Q^2)$
Azimi's update equation for $\hat{d}$	$Q.Q+3Q$	$O(Q^2)$

TABLE IV
RUNTIME COMPLEXITY OF THE DIFFERENT VERSIONS OF THE FORWARD PROCEDURE WITH ORDER OF THE FUNCTION (BIG O NOTATION).

VII. COMPUTATIONAL COMPLEXITY OF THE FORWARD PROCEDURE

Table IV gives the computational complexity of the 4 versions of the forward procedure that are discussed in this paper. We denote the number of states with Q and the maximal length of the duration vectors (describing the state's duration distribution) with D. Azimi's and our approximation give a similar runtime performance as HMMs, while with our assumptions, Yu's equations can be sped up with a factor D.

VIII. CONCLUSION

We showed that simplifying the forward procedure for HSMMs by only considering the expected state duration only gives a satisfiable approximation for left-2-right HSMMs. HSMMs relax the assumption that the hidden state variable captures all state information by also considering state duration. State changes are also determined by state durations. They help to predict state changes in cases that observations do not provide all information. State duration is therefore added to the forward procedure. But this makes the procedure even more compute intensive. The simplification of only considering expected state duration showed, however, to be problematic. The update equation proposed by Azimi et al. [1] was showed to be wrong. But even an improved update equation only produces a reasonable approximation for left-2-right HSMMs.

APPENDIX A
PROOF OF THE EXPECTED DURATION EQUATION

The expectation of the definition of $\hat{d}_t(i)$ (Eq. 14) is given by the following weighted average:

$$\hat{d}_t(i) = \sum_{d=1}^{t} d.P(s_{[t-d:t-1} = i|s_t = i, o_{1:t}) \qquad (21)$$

T	Q	HSMM parameter ranges b_i		ϕ_i		a_{ij}^0	Errors General HSMM $\Delta\bar{\alpha}_i$		$\Delta\hat{d}_i$		Errors left-2-right HSMM $\Delta\bar{\alpha}_i$		$\Delta\hat{d}_i$	
		μ	σ	μ	σ		avg	max	avg	max	avg	max	avg	max
500	5	[0, 100]	[10,50]	[50, 200]	[10, 50]	[0,1]	0.0017	0.094	0.59	3.07	0.00028	0.064	0.29	1.38
500	5	[45, 55]	[10,50]	[50, 200]	[10, 50]	[0,1]	0.0073	0.25	1.86	24.8	0.0033	0.24	0.97	5.45
500	5	[48, 52]	[10,50]	[50, 200]	[10, 50]	[0,1]	0.013	0.30	3.86	38.0	0.0049	0.29	1.26	6.81
500	5	[48, 52]	[10,50]	[50, 200]	[50, 200]	[0,1]	0.016	0.30	4.47	41.1	0.007	0.237	2.04	12.5

TABLE III
EXPERIMENTAL RESULTS ON THE APPROXIMATION ERRORS WITH RANDOMLY-GENERATED HSMMS

We start with writing the definition for $\hat{d}_{t+1}(i)$:

$$\hat{d}_{t+1}(i) = \sum_{d=1}^{t+1} d.P(s_{[t+1-d:t} = i | s_{t+1} = i, o_{1:t+1}) \quad (22)$$

$$= \sum_{d=2}^{t+1} d.P(s_{[t+1-d:t} = i | s_{t+1} = i, o_{1:t+1}) + P(s_t = i, s_{t-1} \neq i | s_{t+1} = i, o_{1:t+1}) \quad (23)$$

Now, for the probability in the sum we get:

$$P(s_{[t+1-d:t} = i | s_{t+1} = i, o_{1:t+1}) = P(s_t = i | s_{t+1} = i, o_{1:t+1}) \cdot P(s_{[t+1-d:t-1} = i | s_t = s_{t+1} = i, o_{1:t+1}) \quad (24)$$

In the second factor we ommit the information about the current state and the current observation:

$$P(s_{[t+1-d:t-1} = i | s_t = s_{t+1} = i, o_{1:t+1}) \approx P(s_{[t+1-d:t-1} = i | s_t = i, o_{1:t}) \quad (25)$$

An approximation which holds under the following independencies

$$s_{t+1} \perp\!\!\!\perp s_{[t+1-d:t-1} | s_t, o_{1:t} \quad (26)$$

$$o_{t+1} \perp\!\!\!\perp s_{[t+1-d:t-1} | s_t, o_{1:t} \quad (27)$$

where $A \perp\!\!\!\perp B|C$ indicates that A and B are independent when conditioned on C. These independencies hold for HMMs (even without conditioning on $o_{1:t}$), but they do not hold for HSMMs since the state duration (expressed by $s_{[t+1-d:t-1}$) determines the system evolution. On the other hand, state duration is partially known by the observations, $o_{1:t}$. Thus, the approximation is reasonable as long as the uncertainty on the states is within limits.

When applying the approximation on Eq. 23 (also on the second term) we find a practical induction step by replacing d with $d' + 1$:

$$\hat{d}_{t+1}(i) = P(s_t = i | s_{t+1} = i, o_{1:t+1}) \quad (28)$$

$$\cdot \left(\sum_{d=2}^{t+1} d \cdot P(s_{[t+1-d:t-1} = i | s_t = i, o_{1:t}) + P(s_{t-1} \neq i | s_t = i, o_{1:t})\right) \quad (29)$$

$$= P(s_t = i | s_{t+1} = i, o_{1:t+1}) \cdot \left(\sum_{d'=1}^{t} (d'+1) P(s_{[t-d':t-1} = i | s_t = i, o_{1:t}) + P(s_{t-1} \neq i | s_t = i, o_{1:t})\right) \quad (30)$$

$$= P(s_t = i | s_{t+1} = i, o_{1:t+1})(\hat{d}_t(i) + 1) \quad (31)$$

Note that the last derivation is based on $\sum_{d'=1}^{t} P(s_{[t-d':t-1} = i | s_t = i, o_{1:t}) + P(s_{t-1} \neq i | s_t = i, o_{1:t}) = 1$

APPENDIX B
PRACTICAL CALCULATION OF THE INDUCTION EQUATION

The induction equation for the expected duration (Eq. 17) is based on $P(s_t = i | s_{t+1} = i, o_{1:t+1})$. We now derive the equation to calculate this probability based on the forward variable and model parameters.

$$P(s_t = i | s_{t+1} = i, o_{1:t+1}) = P(s_t = s_{t+1} = i | o_{1:t+1}) / P(s_{t+1} = i | o_{1:t+1}) \quad (32)$$

Then, due to $o_{t+1} \perp\!\!\!\perp o_{1:t} | s_t = s_{t+1}$[1] and, next, $o_{t+1} \perp\!\!\!\perp s_t | s_{t+1}$, we can write:

$$P(s_t = s_{t+1} = i | o_{1:t+1}) = P(s_t = s_{t+1} = i | o_{1:t}) \cdot P(o_{t+1} | s_t = s_{t+1} = i) / P(o_{t+1} | o_{1:t}) \quad (33)$$

$$= P(s_t = i | o_{1:t}) P(s_{t+1} = i | s_t = i, o_{1:t}) \cdot P(o_{t+1} | s_{t+1} = i) P(o_{1:t}) / P(o_{1:t+1}) \quad (34)$$

Then, we approximated the state transition calculation by considering the expected duration (Eq. 15):

$$P(s_{t+1} = i | s_t = i, o_{1:t}) = \sum_{d_t} a_{ii}(d_t) P(d_t | o_{1:t}) \quad (35)$$

$$\approx a_{ii}(\hat{d}_t(i)) \quad (36)$$

Based on the HSMM parameters, Eq. 32 becomes

$$P(s_t = i | s_{t+1} = i, o_{1:t+1}) = \frac{P(s_{t+1} = i | s_t = i, o_{1:t}) P(s_t = i | o_{1:t}) P(o_{1:t}) P(o_{t+1} | s_{t+1} = i)}{P(s_{t+1} = i | o_{1:t+1}) P(o_{1:t+1})} = \frac{\bar{\alpha}_t(i) a_{ii}(\hat{d}_t(i))}{\bar{\alpha}_{t+1}(i)} \cdot \frac{b_i(o_{t+1})}{P(o_{t+1} | o_{1:t})}$$

where $P(o_{t+1}|o_{1:t})$ is the normalization factor that comes from calculating $\bar{\alpha}$ (Sec. IV).

REFERENCES

[1] M. Azimi, P. Nasiopoulos, and R. Ward, "Offline and online identification of hidden semi-Markov models," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2658 – 2663, 2005.

[2] S.-Z. Yu, "Hidden semi-Markov models," *Artificial Intelligence*, vol. 174, pp. 215–243, 2010.

[3] R. Donat, L. Bouillaut, P. Aknin, P. Leray, and D. Levy, "A generic approach to model complex system reliability using graphical duration models," in *Procs of the Fifth International Mathematical Methods in Reliability Conference*, 2007.

[4] 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, 1989.

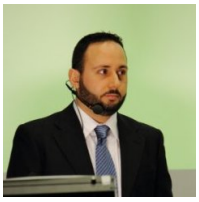
[5] M. Azimi, P. Nasiopoulos, and R. Ward, "Online identification of hidden semi-Markov models," in *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2003.

[6] S.-Z. Yu and H. Kobayashi, "Practical implementation of an efficient forwardbackward algorithm for an explicit-duration hidden Markov model," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1947–1951, 2006.

[7] B. C. Li and S. Z. Yu, "A robust scaling approach for implementation of HSMMs," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1264–1268, 2015.

[1] if $B \perp\!\!\!\perp C|A$, then $P(A|B,C) = P(A|B)P(C|A)/P(C|B)$

Jan Lemeire Jan Lemeire is professor at the Department of Industrial Sciences (INDI) and the Department of Electronics and Informatics (ETRO) at the Vrije Universiteit Brussel (VUB). He is the author of several publications in top journals on parallel computing and probabilistic graphical models. Within the field of parallel computing, his areas of expertise are MPI, GPU computing, performance analysis and performance models. His interests in the field of probabilistic graphical models include causal analysis and learning algorithms in the context of modeling static and dynamic systems.



Francesco Cartella Francesco Cartella is working as Data Scientist at Accenture. He is still an affiliate Member of the Department of Electronics and Informatics (ETRO) of the Vrije Universiteit Brussel (VUB). He is doing research on dynamic models for prognostics for optimization of maintenance of industrial machines. He achieved his PhD entitled 'Hidden semi-Markov Models for online failure prediction of industrial machines' in February 2015.