International Mathematical Forum, Vol. 7, 2012, no. 58, 2871 - 2883

Extended Viterbi Algorithm for Hidden Markov Process: A Transient/Steady Probabilities Approach

Reza A. Soltan and Mehdi Ahmadian[1]

CVeSS, Mechanical Engineering Department, Virginia Tech
Blacksburg, VA 24060, USA

[1]Corresponding Author, e-mail: ahmadian@vt.edu

Abstract

In this paper an extended Viterbi algorithm is presented for first-order hidden Markov processes, with the help of a dummy combined state sequence. For this, the Markov switching's transient probabilities and steady probabilities are studied separately. The algorithm gives a maximum likelihood estimate for the state sequence of a hidden Markov process. Comparing with the standard Viterbi algorithm, this method gives a higher maximum likelihood, and also picks up the state switching earlier, which is particularly important for the out of sample applications. The theory of this method is discussed in this paper and then a sample of a series of experiment is presented to illustrate the theory. A quantitative comparison is also given between this method and the standard Viterbi algorithm.

Keywords: Extended Viterbi; First Order HMM; Combined State Sequence; Hidden Markov model; Maximum Likelihood Estimation; Transient Probability; Steady Probability

I. INTRODUCTION

Hidden Markov model (HMM) has been successfully used by many researches in different branches of science and industry [1-3]. Especially in speech recognition and handwritten script recognition, HMM has been vastly

used [4-10]. A well known solution to this model is the Viterbi algorithm [11, 12]. Through an observation sequence polluted by memoryless noise, Viterbi gives and optimal estimation of the state sequence in a sense of maximum *a posteriori* probabilities [12, 13]. From the inception of HMM till today, various derivations of hidden Markov have been derived and explored. Second-order HMMs (2HMM) [13, 14], hidden semi Markov models (HSMM) [15-17] and Duration-dependent hidden Markov models [8, 18-23] are some examples of these derivations. Russell and Cook in [10, 24] address the property of the underlying model of state duration in the full context of speech pattern modeling using first-order and second-order Viterbi. In that regards they presented an experimental evaluation of two extensions: Hidden semi-Markov models (HSMMs) and extended state HMMs (ESHMMs), where each state of HMM is modeled by a separate sub-HMM that outputs the pdf of the duration of that state [10]. The distributions considered in this research were Poisson and Gamma, and the method was theoretically extended to other distributions.

Mitchell et al. in [18] looked at the complexity of standard Viterbi with explicit duration HMMs. In that regards they introduced a new recursion method that the cost of training is significantly lower than other HMMs with duration modeling.

Burshtein in [25] introduced a robust parametric modeling of durations in HMMs. In that work he proposed a modified Viterbi algorithm in speech recognition in such a way that by incorporating both state and word duration modeling, the error rate could be reduced by 29-43%.

Recently, in 2002 Djuric et al. in [26] introduced an MCMC sampling approach for estimation of non-stationary HMMs. In that regards, they also considered a time dependent transition probability structure that indirectly models Viterbi decoding by a probability mass function. More recently in 2005, Johnson, [23] addressed the capacity and complexity of HMM duration modeling techniques via different versions of Viterbi decoding methods. Johnson studied the standard and extended HMM methods with specific duration-focused approach.

Lately, during the last three years, from 2009 to the current date, several works have been done to present standard and extended Viterbi algorithms, with and without duration dependency [14, 15, 17, 21, 27-29]. Also we would like to address two valuable literature review in this branch of science, one in 1996 by Ostendorf et al. [30], and another one which is a unified review of more recent works in 2010 by Yu [16].

In this paper, we take a different approach in order to improve the extended Viterbi algorithm. In this regards, we explain the probabilities of Markov switching with two separate definitions: Transient Probabilities, and Steady Probabilities. To our knowledge this is the first time that the transient and steady responses of probability estimations of HMM are studied with a 2-step HMM. Therefore we introduce a two-step HMM, where the first step uses a standard Viterbi. In the second step, however, we introduce a combined state

sequence and use an extended Viterbi to model the transient and steady probabilities.

The rest of the paper is organized as follows: Section 2 introduces the first step. Section 3 introduces the second step of our derivation. Section 4 studies an experimental example to illustrate the theory and section 5 conducts a comparison with previous methods. Finally section 6 concludes the paper.

II. FIRST STEP: STANDARD VITERBI

Let $\mathbf{X} = (x_1, x_2, \ldots, x_k)$ represent an N state, k time-long Markov process, where x_t $(1 \le t \le k)$ could be any of the N states. Let $\mathbf{Z} = (z_1, z_2, \ldots, z_k)$ represent an observation sequence, where z_t $(1 \le t \le k)$ is a discretized measure of a continuous time observation universe. A basic assumption of HMM holds where the observation is memoryless, i.e. for any t, the observation z_t depends only on the current state x_t.

Now, lets introduce the notations that we will be using throughout this paper:

- $\Pr(\mathbf{X})$: Total probability of the state sequence $\mathbf{X}$;
- $\Pr(\mathbf{X},\mathbf{Z})$: Joint probability of state sequence $\mathbf{X}$ and the observation sequence $\mathbf{Z}$;
- $\Pr(\mathbf{X}|\mathbf{Z})$: Probability of state sequence $\mathbf{X}$, conditional on the observation sequence $\mathbf{Z}$;
- $\Pr(\mathbf{Z}|\mathbf{X})$: Probability of observation sequence $\mathbf{Z}$, conditional on the state sequence $\mathbf{X}$;
- $\Pr(x_1)$: Initial state probability;
- $\Pr(x_{t+1} \mid x_t)$: State transition probability;
- $\Pr(z_t \mid x_t)$: Probability that z_t is observed at time t, given the state x_t at the same time;

Our aim in the first step is to find a particular state sequence $\mathbf{X}^*$, when observation sequence $\mathbf{Z}$ is given, so that $\Pr(\mathbf{X}^*|\mathbf{Z})$ is maximized. This is also equivalent to $\Pr(\mathbf{X}^*,\mathbf{Z}) = \Pr(\mathbf{X}^*|\mathbf{Z}) \cdot \Pr(\mathbf{Z})$ being maximized. The solution to this problem via a standard Viterbi is straightforward. We have:

$$\Pr(\mathbf{X},\mathbf{Z}) = \Pr(\mathbf{X}).\Pr(\mathbf{Z}\,|\,\mathbf{X})$$

$$= \Pr(x_1).\Pr(z_1\,|\,x_1).\Pr(x_2\,|\,x_1).\Pr(z_2\,|\,x_2)\ \ldots \tag{1}$$

$$= \Pr(x_1).\Pr(z_1\,|\,x_1).\prod_{t=2}^{k}\Pr(x_t\,|\,x_{t-1}).\Pr(z_t\,|\,x_t)$$

Here the details of finding $\mathbf{X}^* = (x_1^*, x_2^*, \ldots, x_k^*)$ is omitted and could be found in [4].

The main contribution of this work is to find <u>another</u> state sequence that has larger maximum likelihood than the state sequence $\mathbf{X}^*$, that we have found in this section.

III. SECOND STEP: EXTENDED VITERBI

Lets introduce a combined state sequence $\mathbf{\Delta X} = (\Delta x_1, \Delta x_2, \ldots, \Delta x_k)$, where:

$$\Delta x_t = x_t x_{t+1} \qquad 1 \le t \le k-1 \tag{2}$$

Note that $\mathbf{\Delta X}$ is an N^2 state, (k-1) time-long first-order HMM. For example in the speech recognition sense, for the word "*seems*":

$\mathbf{X} = (x_1, x_2, x_3, x_4, x_5)$ = (s, e, e, m, s)
$\mathbf{\Delta X} = (\Delta x_1, \Delta x_2, \Delta x_3, \Delta x_4)$ = (se, ee, em, ms)

However, the observation sequence for the combined state HMM is $\mathbf{Y} = (y_1, y_2, \ldots, y_{k-1})$, where:

$$y_t = \Pr(x_{t+1}^* \mid \mathbf{Z}) - \Pr(x_t^* \mid \mathbf{Z}) \qquad 1 \le t \le k-1 \tag{3}$$

Note that the two elements of the RHS of the Eq. (3) are already estimated from the Eq. (1) of the standard HMM in the previous section.

The objective of the second step HMM in this section is to find the particular state sequence $\mathbf{\Delta X}^*$ so that $\Pr(\mathbf{\Delta X}^*|\mathbf{Y})$, or equivalently, $\Pr(\mathbf{\Delta X}^*,\mathbf{Y}) = \Pr(\mathbf{\Delta X}^*|\mathbf{Y}) \,.\, \Pr(\mathbf{Y})$ is maximized. We have:

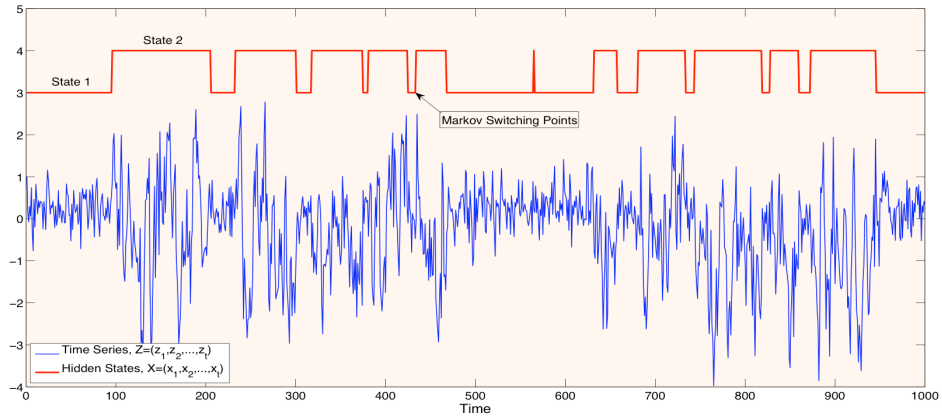

Fig. 1 A N=2 state HMM sample with its associated time series

$$\begin{aligned}\Pr(\Delta\mathbf{X},\mathbf{Y}) &= \Pr(\Delta\mathbf{X}).\Pr(\mathbf{Y}\mid\Delta\mathbf{X})\\ &= \Pr(\Delta x_1).\Pr(y_1\mid\Delta x_1).\Pr(\Delta x_2\mid\Delta x_1).\Pr(\Delta y_2\mid\Delta x_2)\ \ldots\\ &= \Pr(\Delta x_1).\Pr(y_1\mid\Delta x_1).\prod_{t=2}^{k}\Pr(\Delta x_t\mid\Delta x_{t-1}).\Pr(y_t\mid\Delta x_t)\end{aligned} \tag{4}$$

Lets denote Δx_t^* by $x_t^* x_{t+1}^*$. Although the states for this HMM may look like a second-order HMM, they are not. Note that each combination of Δx belongs to a separate state, since the second step HMM has N^2 states. In other words each Δx_t can be considered as a single state, associated with a N^2 state, (k-1) time-long first-order HMM.

Steady vs. Transient Probabilities

Δx explains the steady state versus transient probabilities. If in the notation $x_{t-1}^* x_t^*$ both consecutive parts, i.e. x_{t-1}^* and x_t^* are the same, it means that there is more likelihood that the Markov state at the next time step stays on the same state. On the other hand where the two consecutive parts of $x_{t-1}^* x_t^*$ are not the same means that the Markov switching is on the transient from x_{t-1}^* to x_t^*. In the speech recognition sense, again for the word "*seems*":

$\mathbf{\Delta X} = (\Delta x_1, \Delta x_2, \Delta x_3, \Delta x_4)$ = (se, ee, em, ms)

shows that the second state is a steady state of state "e", and every other state is transient.

This notion of transient could be also seen in a way that N^2 states of the combined state sequences, can be written on an $N\times N$ matrix where the diagonal

elements are steady probabilities and off-diagonal elements correspond to transient probabilities.

Introducing the steady and transient states helps us know that; when the state goes to transient response, it will soon end up on the second state of the two combined states.

Therefore we introduce our final single state sequence, which is always the second part of the combined state sequence. In this way if the combined state sequence is steady, then the second part of it is in fact the current state. If the combined state is transient then the second part is the state that the transition is going to end up to. Thus this state sequence estimation is in general ahead of time.

Note that, again finding the $\mathbf{\Delta X}^*$ is the solution to a separate standard Viterbi [4] and is straightforward and is omitted here.

Now we introduce a single state sequence $\hat{\mathbf{X}} = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_k)$, so that:

$$\begin{aligned} &\hat{x}_{t+1} = x_{i+1} \quad \text{where:} \\ &x_{i+1} = \text{Argmax}\{\Pr(x_i^* x_{i+1}^* \mid \mathbf{Y})\} \quad 1 \le i \le k-1 \\ &\text{and } \hat{x}_1 = x_1 \end{aligned} \tag{5}$$

Note that $\Pr(\Delta x_t^* | \mathbf{Y}) = \Pr(x_t^* x_{t+1}^* | \mathbf{Y})$ is already estimated as the output of our second step HMM via maximization of Eq. (4).

The joint probability of states and observation sequence for the second step HMM could be noted as follows:

$$\begin{aligned} &\Pr(\hat{\mathbf{X}}, \mathbf{Y}) = \\ &= \Pr(\hat{x}_1).\Pr(\hat{x}_2 \mid \hat{x}_1).\Pr(y_1 \mid \hat{x}_2) \ ... \\ &= \Pr(\hat{x}_1).\prod_{t=2}^{k} \Pr(\hat{x}_t \mid \hat{x}_{t-1}).\Pr(y_{t-1} \mid \hat{x}_t) \end{aligned} \tag{6}$$

Adding a y_0 with $\Pr(y_0)=1$ to the observation sequence of Eq. (3), we can linearly transform $\mathbf{Y}= (y_0, y_1, \ldots, y_{k-1})$, to $\hat{\mathbf{Y}} = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_k)$ by shifting the time index forward by 1. So we have:

$$\begin{aligned} &\Pr(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) = \\ &= \Pr(\hat{x}_1).\prod_{t=2}^{k} \Pr(\hat{x}_t \mid \hat{x}_{t-1}).\Pr(\hat{y}_t \mid \hat{x}_t) \end{aligned} \tag{7}$$

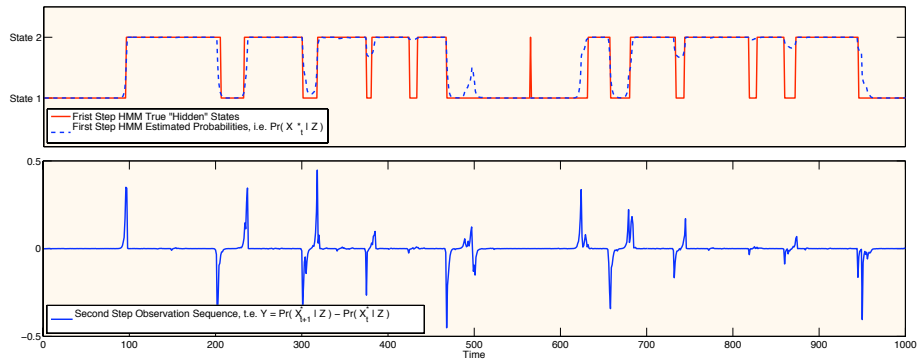

Fig. 2 Top: The true N=2 state sequence and the first step estimation of Markov probabilities

Fig. 2 Bottom: The second step observation sequence, calculated as the difference of the first step Markov probabilities

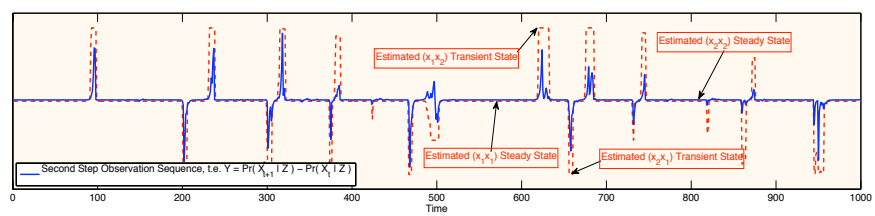

Fig 3. The second step estimated N^2=4 states, for the second step observation sequence

Note that the state sequence of $\hat{\mathbf{X}} = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_k)$ is already calculated from Eq. (5), and will not be estimated from maximizing Eq. (6), nor Eq. (7). However, from the first line of the Eq. (5), we can see that the characteristics of the states $\hat{x}$ are the same as states x. However, the sequence of $\hat{\mathbf{X}}$ is different than the sequence of $\mathbf{X}$. Therefore, the transition probabilities of the two states are the observation probabilities are different. In other words:

$$\begin{aligned} &\hat{x} \equiv x \\ &\Pr(\hat{x}_t \mid \hat{x}_{t-1}) \neq \Pr(x_t \mid x_{t-1}) \\ &\Pr(y_t \mid \hat{x}_t) \neq \Pr(z_t \mid x_t) \end{aligned} \tag{8}$$

Therefore there are two differences between the standard Viterbi state sequence estimation of Eq. (1) and the extended Viterbi state sequence estimation of Eq. (7): First, the observation sequence and probabilities and the transition probabilities are different. Second, the length of Eq. (7) is less than Eq.

(1) by 1 probability measure. This is due to the fact that the first state at Eq. (7) doesn't have an observation associated with it.

Because of the differences between the transition probabilities and the observation probabilities, we cannot directly compare the total maximized likelihood of our extended Viterbi at Eq. (7) with the standard Viterbi of Eq. (1). However, there is a chance that Eq. (7) is greater than Eq. (1) because of its shorter length. Note that the probabilities are less than 1 positive values, so shorter sequences have larger total likelihoods. To compare these two we conducted a series of experiments, which we will talk about them in the next two sections.

IV. AN EXPERIMENT

To illustrate the theory as well as comparing this extended Viterbi algorithm with two step first-order HMMs, we conduct a series of experiments. The result of one of the experiments is given in this section. Figure 1 shows a 2 state (i.e. N=2) HMM with its associated time series of length k=1000[s]. Note that this time series is generated with an AR(1) difference model with different set of parameters at each state. More

specifically the generative models at both states of Figure 1 are:

$$\begin{cases} z_{t+1} = +0.2 + 0.1 z_t + \varepsilon_1 \ : \varepsilon_1 \sim N(0,0.2), \ \text{if } x_t=1 \\ z_{t+1} = -0.2 + 0.7 z_t + \varepsilon_1 \ : \varepsilon_1 \sim N(0,0.9), \ \text{if } x_t=2 \end{cases} \tag{9}$$

$$\text{where: } z_1 = 0$$

The sequence of hidden states, i.e. **X** is shown in the top of the Figure 1. **X** is essentially what we are trying to estimate, having **Z** as observation sequence. Top exhibit of the Figure 2, also shows the hidden state sequence **X**, and its estimate based on a standard one-step first-order HMM. Our goal in this research was to find a state sequence that has less error and higher total likelihood than this.

Bottom exhibit of Figure 2 shows the observation sequence for the second step HMM of our derivation. This is essentially calculated from the estimated probabilities of the first step HMM via Eq. (3). The nest step is the second step HMM that we run for this problem. Here we conduct a N^2=4 step HMM on the observation sequence of the Bottom exhibit of figure 2.

Figure 3 shows the estimated four states of the second step of our HMM via Eq. (4). Note that the hidden states for the second step are defined by Eq. (2).

After the second step is complete from the, from the dotter red curve of the Figure 3, we can calculate our extended Viterbi solution $\hat{\mathbf{X}} = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_k)$ via Eq. (5). Top exhibit of the Figure 4, shows both the first step and the second step state probability estimations. Note that first step (i.e. Blue dotted curve) is the

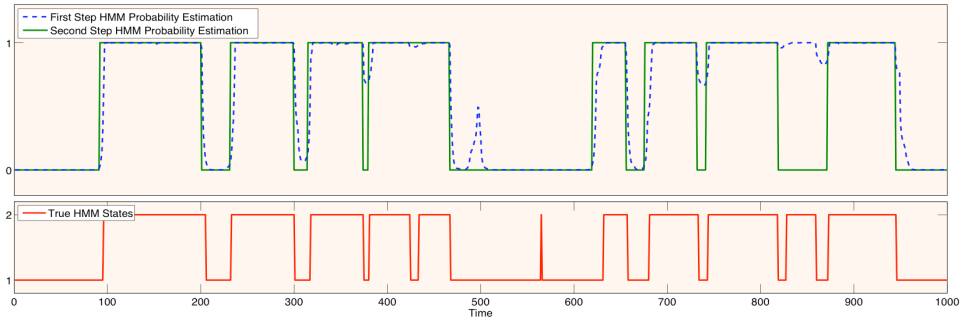

Fig. 4 Top: The first step versus second step state probabilities estimation
Fig. 4 Bottom: The true state sequence is shown for comparison

standard Viterbi solution via a standard first orders HMM, and green solid curve is the extended Viterbi solution via a two-step HMM.

It's good to emphasis here that in the top exhibit of the figure 4, the green solid curve is still the estimated probabilities, not the state sequences, where the bottom exhibit of figure 4 is the state sequence. The fact that the second step probabilities are a lot cleaner then the first step estimation, shouldn't make us confused. This bottom exhibit of the figure 4 shows the state sequence that we were trying to estimate. This figure clearly shows the advantages of our state estimation method versus the standard Viterbi.

V. A COMPARISON

The extended Viterbi algorithm via a two-step HMM that we discussed in this paper has one major advantage and one major weakness comparing to the standard Viterbi:

Weakness:

The computational complexity of our extended Viterbi is larger than the standard Viterbi. This is due to two facts: Firstly, the extended Viterbi runs the HMM algorithm twice. And Secondly, the second step of our extended Viterbi runs on a N^2 state HMM. In general the computational complexity of the standard Viterbi (1-step HMM) versus the extended Viterbi (2-step HMM) is $O\{kN^2\}$ versus $O\{k(N+N^2)^2\}$ respectively, where k is the length of time series.

This difference in the computational complexity is especially noticeable for larger size HMMs.

Advantage:

The advantage of this method versus the standard Viterbi is the greater accuracy of estimation. In other words as we can see in Top exhibit of Figure 4, the state estimation of the extended Viterbi is more accurate and cleaner than the standard Viterbi. Furthermore there are multiple times that the standard Viterbi with a one step HMM, misses the Markov switching points. This is shown at time $t=380^s$, $t=740^s$, and $t=830^s$, where the 2-step HMM catches the Markov switching but the 1-step is not able to catch it and skips it.

Basically with this method we are sacrificing some computational efficiency to get a more accurate state estimation.

To quantify the improvement of this method over the standard Viterbi, a series of similar experiments (30 experiments with 2 and 3 state HMMs, with time series of the length $k=1000^s$) have been conducted and the differences between the outputs of the two methods have been observed. On average this extended Viterbi method gives 29% improvement over the standard Viterbi. Along the series of experiments, a range of 14% to 41% improvement over the standard Viterbi was observed.

The particular experiment that was shown and studied in the previous section showed 21% improvement of state estimation over the standard Viterbi algorithm.

VI. CONCLUSION

An extended Viterbi algorithm is presented for first-order hidden Markov processes, with the help of a dummy combined state sequence. For this, the Markov switching's transient probabilities and steady probabilities are studied separately. The algorithm gives a maximum likelihood estimate for the state sequence of a hidden Markov process. Comparing to the standard Viterbi algorithm, this method gives a higher maximum likelihood, and also picks up the state switching earlier, which is particularly important for the out of sample applications. The theory of this method was discussed in this paper and then an experiment was presented to illustrate the theory. A comparison was also given between this method and the standard Viterbi algorithm.

References

[1] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. New York: Springer-Verlag, 1995.

[2] A. B. Poritz, "Hidden Markov models: a guided tour," *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing,* pp. 7-13, 1988.

[3] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition," *The Bell System technical journal,* vol. 62, p. 1035, 1983.

[4] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE,* vol. 77, pp. 257-286, 1989.

[5] J. K. Baker, "Trainable grammars for speech recognition," *The Journal of the Acoustical Society of America,* vol. 65, p. S132, 1979.

[6] J. Baker, "Stochastic modeling for automatic speech understanding," *Speech Recognition (D. R. Reddy editor), Academic Press, New York,* 1975.

[7] F. Jelinek, "Statistical Methods for Speech Recognition," *The Press (Christchurch, N.Z.),* 1998.

[8] P. Ramesh and J. G. Wilpon, "Modeling state durations in hidden Markov models for automatic speech recognition," in *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, 1992, pp. 381-384 vol.1.

[9] L. R. Bahl, "A Maximum Likelihood Approach to Continuous Speech Recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. PAMI-5, pp. 179-190, 1983.

[10] M. Russell and A. Cook, "Experimental evaluation of duration modelling techniques for automatic speech recognition," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87.*, 1987, pp. 2376-2379.

[11] G. D. Forney Jr, "The viterbi algorithm," *Proceedings of the IEEE,* vol. 61, p. 268, 1973.

[12] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE transactions on information theory,* p. 260, 1967.

[13] Y. He, "Extended Viterbi algorithm for second order hidden Markov process," in *Pattern Recognition, 1988., 9th International Conference on*, 1988, pp. 718-720 vol.2.

[14] X. Yi, Y. Shun-zheng, T. Shensheng, and H. Xiangnong, "A Two-Layer Hidden Markov Model for the Arrival Process of Web Traffic," in *Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on*, 2011, pp. 469-471.

[15] C. Wei-ho and Y. Kung, "Modified hidden semi-markov model for modelling the flat fading channel," *Communications, IEEE Transactions on,* vol. 57, pp. 1806-1814, 2009.

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

[17] M. Dong and Y. Peng, "Equipment PHM using non-stationary segmental hidden semi-Markov model," *Robot. Comput.-Integr. Manuf.,* vol. 27, pp. 581-590, 2011.

[18] C. Mitchell, M. Harper, and L. Jamieson, "On the complexity of explicit duration HMM's," *Speech and Audio Processing, IEEE Transactions on,* vol. 3, pp. 213-217, 1995.

[19] Y. K. Park, C. K. Un, and O. W. Kwon, "Modeling acoustic transitions in speech by modified hidden Markov models with state duration and state duration-dependent observation probabilities," *Speech and Audio Processing, IEEE Transactions on,* vol. 4, pp. 389-392, 1996.

[20] N. B. Yoma and J. S. Sanchez, "MAP speaker adaptation of state duration distributions for speech recognition," *Speech and Audio Processing, IEEE Transactions on,* vol. 10, pp. 443-450, 2002.

[21] S. Winters-Hilt, Z. Jiang, and C. Baribault, "Hidden Markov model with duration side information for novel HMMD derivation, with application to eukaryotic gene finding," *EURASIP J. Adv. Signal Process,* vol. 2010, pp. 1-11, 2010.

[22] A. Anastasakos, R. Schwartz, and S. Han, "Duration modeling in large vocabulary speech recognition," in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, 1995, pp. 628-631 vol.1.

[23] M. T. Johnson, "Capacity and complexity of HMM duration modeling techniques," *Signal Processing Letters, IEEE,* vol. 12, pp. 407-410, 2005.

[24] M. Russell and R. Moore, "Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85.*, 1985, pp. 5-8.

[25] D. Burshtein, "Robust parametric modeling of durations in hidden Markov models," in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, 1995, pp. 548-551 vol.1.

[26] P. M. Djuric and C. Joon-Hwa, "An MCMC sampling approach to estimation of nonstationary hidden Markov models," *Signal Processing, IEEE Transactions on,* vol. 50, pp. 1113-1123, 2002.

[27] S. Winters-Hilt and J. Zuliang, "A Hidden Markov Model With Binned Duration Algorithm," *Signal Processing, IEEE Transactions on,* vol. 58, pp. 948-952, 2010.

[28] L. M. Lee, "High-Order Hidden Markov Model and Application to Continuous Mandarin Digit Recognition " *Journal of Information Science And Engineering,* vol. 27, pp. 1919-1930, 2011.

[29] S. Calinon, A. Pistillo, and D. G. Caldwell, "Encoding the time and space constraints of a task in explicit-duration Hidden Markov Model," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, pp. 3413-3418.

[30] M. Ostendorf, V. V. Digalakis, and O. A. Kimball, "From HMM's to segment models: a unified view of stochastic modeling for speech recognition," *Speech and Audio Processing, IEEE Transactions on,* vol. 4, pp. 360-378, 1996.

Received: June, 2012