Information Processing Letters 112 (2012) 732–737

Contents lists available at SciVerse ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

An efficient algorithm for parameterizing HsMM with Gaussian and Gamma distributions

Y. Xie [a,*], S. Tang [b], C. Tang [c], X. Huang [d]

[a] *School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510275, China*
[b] *Department of Engineering Technology, Missouri Western State University St. Joseph, MO 64507, USA*
[c] *School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin 541004, China*
[d] *Network and Information Technology Center, Sun Yat-Sen University, Guangzhou 510275, China*

ARTICLE INFO

Article history:
Received 14 December 2011
Received in revised form 12 June 2012
Accepted 13 June 2012
Available online 18 June 2012
Communicated by Jinhui Xu

Keywords:
Algorithm
Parametric HsMM
Gaussian
Gamma

ABSTRACT

A widely used method for parameterizing hidden semi-Markov model is using Gaussian distribution to form the output probability and using Gamma distribution to form the state duration probability. Most of these models are based on the classical Newton's method with second-order convergence, whose iterative convergence speed is slow for large-scale realtime applications. An improved parameter re-estimation algorithm is introduced for such parametric hidden semi-Markov model in this paper. The proposed approach is based on forward and backward algorithm. It applies an iterative method with eighth-order convergence to improve the performance of the model. The numerical examples validate the proposed method.

1. Introduction

A hidden Markov model (HMM) [1] is defined as a doubly stochastic process. The underlying stochastic process is a discrete-time finite-state homogeneous Markov chain. The state sequence is not observable and controls the observable process. The HMMs are an important class of models that are successful in many application areas. However, the state duration of an HMM is implicitly a geometric distribution or exponentially distributed according $p(d) = (1 - a_{ii})a_{ii}^{d-1}$. This makes the HMM has limitations in some applications. In order to solve this issue, an extension of the HMM named hidden semi-Markov model (HsMM) is proposed. HsMM is defined by allowing the underlying process to be a semi-Markov chain. Each state has a variable duration, which is associated with the number of observations produced while in the state. The HsMM is also called "explicit duration HMM" [1], "variable-duration HMM" [2], segmental HMM [3] and hidden semi-Markov model [4] in the literature.

The first approach to hidden semi-Markov model was proposed by Ferguson [5]. However, Ferguson's algorithm is computationally too expensive to be of practical use in many applications [4]. In order to reduce the computational complexity of HsMM, many approaches, e.g., [6] utilized the probability distributions with continuous probability density functions to form the continuously variable duration and output. Among those previous work, Gaussian distribution and Gamma distribution are the two most commonly used means for parameterizing the output probability and the state duration probability of HsMM. The main advantage of such a parametric HsMM is threefold: (i) A finite mixture of Gaussian components can model/approximate any continuous distribution with arbitrary precision without any assumption with respect to distributional properties of the raw data analyzed [7]. (ii) Gamma distribution is a flexible function to express distributions of different shapes (e.g., exponential/right-skewed/Gaussian distribution) by adjusting its parameters. (iii) Both of them only have two parameters. In [6],

* Corresponding author.
E-mail address: xieyi5@mail.sysu.edu.cn (Y. Xie).

0020-0190/$ – see front matter
http://dx.doi.org/10.1016/j.ipl.2012.06.005

Levinson first discussed the re-estimation algorithm for HsMM whose state durations are Gamma distribution. After that, Levinson's model has been widely used in different areas. However, the main drawback of Levinson's algorithm is two-fold: (i) Its re-estimation algorithm is derived from [5] which is computationally expensive. (ii) The classical Newton's method with second-order convergence is applied to solve the parameters of Gamma distribution, whose iterative convergence speed is slow for large-scale realtime applications.

In this paper, we propose an improved algorithm for the widely used parametric HsMM which uses Gaussian distribution and Gamma distribution to form the output probability and state duration probability, respectively. The main contribution of this paper is that the proposed iterative method with eighth-order convergence improve the efficiency of the Gaussian and Gamma-based parametric HsMM (G^2PHsMM) and enables the model to be used in most real-time applications.

2. The definition of HsMM

The basic HsMM consists of a pair of stochastic processes: the observed process $\{\vec{o}_t\}$ and the hidden state process $\{X_t\}$, where $t \in \{1, 2, \ldots\}$ is the number of observation (also called event). $\{\vec{o}_t\}$ is associated with $\{X_t\}$ by the conditional distribution depending on the state process which is a finite-state semi-Markov chain. The conditional distributions usually overlap and so, in general, a specific observation can arise from more than one state. Thus $\{X_t\}$ is not observable directly through $\{\vec{o}_t\}$ but can be estimated. $\{\vec{o}_t\}$ itself may be either discrete or continuous, univariate or multivariate. The basic structure of the HsMM is illustrated in Fig. 1. Some notations and definitions used in this paper are listed in Table 1. We assume that the output process at the tth event depends only on the state of the underlying semi-Markov chain at event t, i.e., $Pr[\vec{O}_t = \vec{o}_t \mid \vec{O}_{1|t-1} = \vec{o}_{1|t-1},\ X_{1|t} = x_{1|t}] = Pr[\vec{O}_t = \vec{o}_t \mid X_t = x_t]$.

3. The improved algorithm for G^2PHsMM

In G^2PHsMM, the representation of the observation PDF, i.e., $b_m(\vec{o}_t)$, is a finite mixture of the form:

$$b_m(\vec{o}_t) = \sum_{k=1}^{\kappa_m} c_{mk} b_{mk}(\vec{o}_t) = \sum_{k=1}^{\kappa_m} c_{mk} \mathcal{N}(\vec{o}_t, \vec{\mu}_{mk}, \Sigma_{mk}) \quad (1)$$

where κ_m is the number of Gaussian component in state m, $\mathcal{N}(\vec{o}_t, \vec{\mu}_{mk}, \Sigma_{mk})$ denotes the multi-dimensional normal density function with mean vector $\vec{\mu}_{mk}$ and covariance matrix Σ_{mk} for the kth component in state m. $c_{mk} = P[Y_{mt} = k \mid X_t = m]$ is the conditional weight for the kth mixture component in state m. c_{mk} satisfies the following stochastic constraints: $c_{mk} \geqslant 0$ and $\sum_{k=1}^{\kappa} c_{mk} = 1$ for $m \in \mathbb{M}$ and $k \in [1, \kappa_m]$. Thus, the $b_m(\vec{o}_t)$ is properly normalized, i.e., $\int_{\mathbb{O}} b_m(\vec{o})\, d\vec{o} = 1$.

For the PDF of state duration, the representation of $p_m(d)$ is given as follows:

$$p_m(d) = \frac{1}{\Gamma(\nu_m)} d^{\nu_m - 1} e^{-\omega_m d} \omega_m^{\nu_m}, \quad \nu_m,\ \omega_m > 0 \quad (2)$$

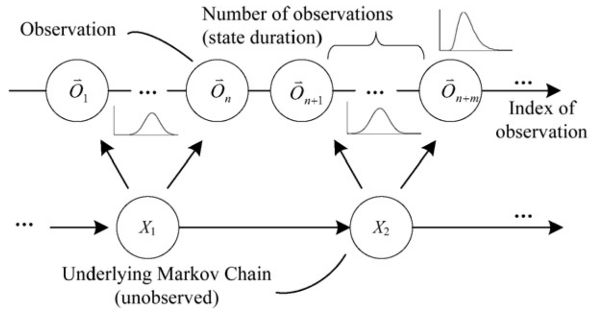

Fig. 1. Structure of HsMM.

Table 1
Notation of symbols.

$\mathbb{O}$	the output space of the model
$\mathbb{M}$	$= \{1, 2, \ldots, M\}$, the state space including M states of a semi-Markov chain
$\mathbb{D}$	$= \{1, 2, \ldots\}$, the possible state duration
O	$= \{\vec{o}_1, \vec{o}_2, \ldots, \vec{o}_T\}$, the observed event series with T observations
$\vec{o}_t$	$\in \mathbb{O}$, the observed vector at time t
$\vec{o}_{a\|b}$	$= \{\vec{o}_a, \vec{o}_{a+1}, \ldots, \vec{o}_b\}$
X	$= \{X_1, X_2, \ldots, X_T\}$, the underlying state sequence
X_t	$\in \mathbb{M}$, the underlying state at time t
τ_t^r	$\in \mathbb{D}$, the remaining time of state X_t
π_m	$\equiv Pr[X_1 = m \mid \lambda]$ the initial state probability
a_{mn}	$\equiv Pr[X_t = n \mid X_{t-1} = m, \lambda]$, the state transition probability
$b_m(\vec{o}_t)$	$\equiv Pr[\vec{O}_t = \vec{o}_t \mid X_t = m, \lambda]$, the observation probabilities
$p_m(d)$	$\equiv Pr[\tau_t^r = d \mid X_t = m, \lambda]$, the state duration probability
λ	$= \{\pi_m, a_{mn}, b_m(\vec{o}_t), p_m(d)\}$, the total parameters of non-parametric HsMM

where $\Gamma(\nu)$ is the Gamma function and can be calculated by $\Gamma(\nu) = (\nu - 1)!$, $\nu \in \mathbb{Z}^+$. The mean value of d is ν_m/ω_m and its variance is ν_m/ω_m^2.

To overcome the shortcomings of existing algorithms designed for G^2PHsMM, a new algorithm is proposed for estimating the parameters of G^2PHsMM (i.e., $\lambda = \{\pi_m, a_{mn}, c_{mk}, \vec{\mu}_{mk}, \Sigma_{mk}, \nu_m, \omega_m\}$). The algorithm is based on forward–backward algorithm [8] and a new iterative method with eighth-order convergence which has been demonstrated valid for solving nonlinear equations [9].

The forward process is defined as:

$$\alpha_t(m, d) \overset{\text{def}}{=} P\big[\vec{o}_{1|t}, \big(X_t, \tau_t^r\big) = (m, d) \,\big|\, \lambda\big]$$
$$= \begin{cases} \pi_m b_m(\vec{o}_1) p_m(d), & t = 1 \\ \alpha_{t-1}(m, d+1) b_m(\vec{o}_t) & \\ \quad + \sum_{n \neq m} \alpha_{t-1}(n, 1) a_{nm} \cdot b_m(\vec{o}_t) p_m(d), & d \geqslant 1 \end{cases} \quad (3)$$

The backward process is defined as:

$$\beta_t(m, d) \overset{\text{def}}{=} P\big[\vec{o}_{t+1|T} \,\big|\, \big(X_t, \tau_t^r\big) = (m, d), \lambda\big]$$
$$= \begin{cases} b_m(\vec{o}_{t+1}) \beta_{t-1}(m, d-1), & d > 1 \\ \sum_{n \neq m} a_{mn} b_n(\vec{o}_{t+1}) & \\ \quad \times \big(\sum_{d \geqslant 1} p_n(d) \beta_{t+1}(n, d)\big), & d = 1 \\ 1, & t = T \end{cases} \quad (4)$$

A joint probability function of observing O and a transition from m to n $(n \neq m)$ at time t given λ is defined by: