Available online at www.sciencedirect.com
SCIENCE DIRECT®

Pattern Recognition 38 (2005) 2022–2031

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

# Hidden Markov models with factored Gaussian mixtures densities

Hao-Zheng Li[a], Zhi-Qiang Liu[b,*], Xiang-Hua Zhu[a]

[a] *Continuing Education School, Beijing University of Posts and Telecommunications, Beijing, China*
[b] *School of Creative Media, City University of Hong Kong, Kowloon, Hong Kong, SAR, China*

Received 23 June 2003; received in revised form 7 December 2004; accepted 17 January 2005

## Abstract

We present a factorial representation of Gaussian mixture models for observation densities in hidden Markov models (HMMs), which uses the factorial learning in the HMM framework. We derive the reestimation formulas for estimating the factorized parameters by the Expectation Maximization (EM) algorithm and propose a novel method for initializing them. To compare the performances of the proposed models with that of the factorial hidden Markov models and HMMs, we have carried out extensive experiments which show that this modelling approach is effective and robust.


*Keywords:* Hidden Markov models; Gaussian mixtures; EM algorithm; Factorial learning

## 1. Introduction

Many real-world observed data are characterized by multiple coupled causes or *factors*. For instance, face images may be generated by combining eyebrows, eyes, nose and mouth. Recently Zemel and Hinton proposed a factorial learning architecture [1,2] to deal with factorial data. The goal of factorial learning is to to discover the multiple underlying causes from the observed data and find a representation that will both parsimoniously describe the data and reflect the underlying causes. Hidden Markov models (HMMs) are probabilistic models that can describe sequential data effectively and have been successfully used in many applications, such as pattern recognition and speech recognition [3]. Williams and Hinton [4] proposed generalizing HMMs through the use of distributed state representations in order to promote the representational efficiency of HMMs. Ghahramani and Jordan [5] studied a special case of this generalization referred to as the factorial hidden Markov models (FHMMs), and considered the use of FHMMs for summarizing sequential data caused by multi-factors. FHMMs have been applied to a simple speech recognition task in Ref. [6] with mixed results. FHMMs use Gaussian density functions as output densities and a common covariance matrix for all the Gaussian density functions. However, the Gaussian distribution assumption of FHMMs does not adequately reflect the nature of real-world data: First, because of their squared exponent, Gaussian density values decay too fast when the observable variable deviates from its mean. Second, variations exist in data; for example, speech signals from multi-speakers, face images from different persons. In such cases, the corresponding probability distributions are generally multimodels. In practice, Gaussian mixture models are used for approximating output data [7,8]. Unfortunately, there is no obvious way of incorporating Gaussian mixtures in the framework of FHMMs [5]. Moreover, the requirement for a common covariance is too restrictive in most applications.

In this paper, we present a new approach that uses factorial learning in the HMM framework. The state with a mixture density in the HMM can be represented in terms

* Corresponding author. Tel.: +852 2194 2832;
fax: +852 2788 7165.
*E-mail address:* smzliu@cityu.edu.hk (Z.-Q. Liu).

0031-3203/$30.00 
doi:10.1016/j.patcog.2005.01.027

of multi-substates with a single mixture density (see Section 3). We factor these substates into multi-substate variables; each single mixture density depends on all the factored multi-substate variables. We introduce a method to combine the mean and covariance matrix of each component in the Gaussian mixture model from different factored parameters. In this way, we can make use of the factored Gaussian mixture models as the output probability density functions (pdfs) for the HMM.

In Section 2, we briefly discuss HMMs and FHMMs. Section 3 describes the factorial representation for Gaussian mixture model for HMMs and presents the parameter estimation process. In Section 4, we present experimental results comparing HMMs with different output density models to show the model quality. In Section 5, we give a general summary and discuss further research directions.

## 2. Hidden Markov models and factorial hidden Markov models

The HMM is a probabilistic model that is able to describe time sequence effectively. Each state in HMM can be thought of as representing particular patterns or regions of sequential data. The probability that an observation has been generated given the model is:

$$P(Y|\lambda) = \sum_S P(S_1)P(Y_1|S_1)\prod_{t=2}^{T} P(S_t|S_{t-1})P(Y_t|S_t), \quad (1)$$

where $T$ is the length of observation sequence (total number of time steps); $Y = \{Y_1, Y_2, \ldots, Y_T\}$ a sequence of possible vector observation; $K$ the number of states; $S = \{S_1, S_2, \ldots, S_T\}$ a sequence of hidden state variables generating $Y$, $S_t$ can take one of $K$ discrete value; $P(S_t = j|S_{t-1} = i)$ is the transition probability from time $t-1$ being state $i$ to time $t$ being state $j$, is specified by a $K \times K$ matrix $A_{ij}$; $P(S_1 = i)$ the probability of being in $i$th state at time $t = 1$, also denoted as $a_i$; $P(Y_t|S_t)$ the pdf of the output vector given the state $S_t$, typically modeled as a mixture of Gaussians; and $\lambda$ the compact notation to indicate the complete parameters.

Fig. 1 shows the conditional independence specified by Eq. (1). The HMM can be viewed as a generalization of mixture models: in addition to the mixture components, the HMM is made up of a series of discrete states and a set of transition probabilities between the states. As for the HMM whose output densities belong to the mixture model distribution at time $t$, the sequential data will stay in state $S_t$ and a group of mixture components with weights attached to this state are active to account for the data at $t$; and at time $t+1$, a new group of active components with weights attached to state $S_{t+1}$ are selected according to the transition probabilities. In contrast, for HMMs whose output densities belong to a single-model distribution, at time $t$, only one mixture component is active to represent the data at $t$.

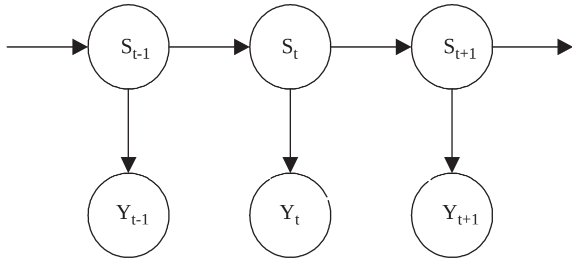


Fig. 1. Dynamic belief network representation of a HMM.

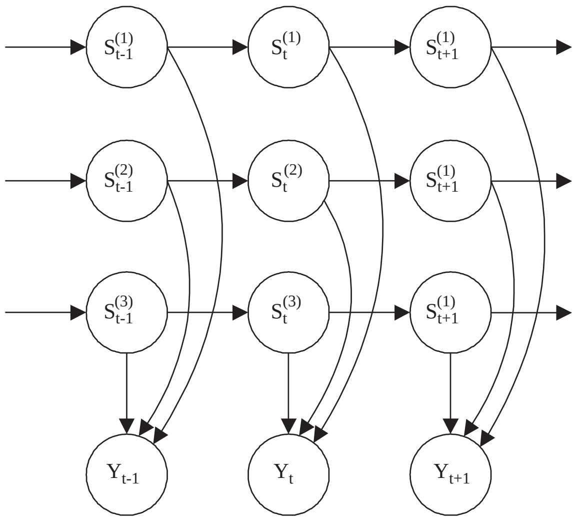


Fig. 2. Dynamic belief network representation of a FHMM with three Markov chains.

The factorial HMM is an extension of the traditional HMM in that it consists of multiple hidden Markov chains as illustrated in Fig. 2. The state of each chain is a single multinomial variable that can take one of $K$ discrete values. Each chain has independent dynamics. The output of the model in each time step is, however, dependent on the values of the state variables of all the chains. In Ref. [5] this is achieved by constructing a *meta-state* composed of $M$ discrete state variables as follows:

$$S_t = S_t^{(1)}, \ldots, S_t^{(m)}, \ldots, S_t^{(M)},$$

where $M$ is the number of chains and $S_t^{(m)}$ is the state of $m$th chain at time $t$. The state space of FHMMs consists of the cross product of these state variables. There are total $K^M$ meta-states representing the time sequence. The output of the meta-state belongs to a Gaussian distribution with the mean being a linear combination of the state means. The transition between the meta-states is the product of the transitions between the states of the same chain. Since the