Factorial HMM

Théis Bazin, Valentin De Bortoli
theis.bazin@ens-paris-saclay.fr, vdeborto@ens-paris-saclay.fr

école normale supérieure paris—saclay

Introduction

We present the Factorial Hidden Markov Model, Factorial HMM, as introduced by Ghahramani and Jordan, an extension of HMM, and methods to perform inference on this model.

Context HMM assumes the hidden variable to be single multinomial variable, this can lead to computational issues,

HMM limitation Suppose we want to describe $M = 30$ binary states [1],

- Using HMM the state space has size 2^{30},
- Using *distributed* reduces this to only $M = 30$ binary states.

This yields the Factorial HMM model:

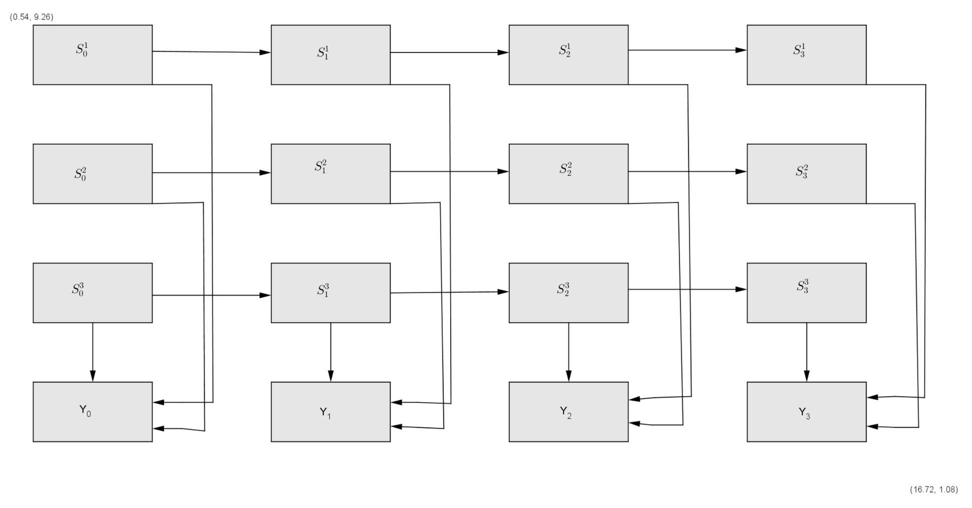

Figure 1: Graphical model for Factorial HMM

An EM algorithm...

Training this model can be done per the EM algorithm [2].

E-step Compute the posterior $p(S_{(t)}^{(m)}|Y_{(t)}, \theta)$

M-step Update the parameters θ

M-step is *easy* but E-step can be *really hard* [1] to compute. Recall that in the HMM we used an alpha-beta recursion to deal with the E-step.
We present three different ways to deal with E-step:

- Exact inference,
- Gibbs sampling,
- Variational methods.

Exact inference

The exact inference allows us to compute the posterior distribution setting:

$$\begin{cases} \alpha_t = p(S_t^1, \ldots, S_t^M, Y_{(1,t)}|\theta) \\ \alpha_t^m = p(S_{t-1}^1, \ldots, S_{t-1}^m, S_t^{m+1}, \ldots, S_t^M|\theta) \\ \beta_t = p(Y_{(t+1,T)}|S_t^1, \ldots, S_t^M, \theta) \\ \beta_{t-1}^m = p(Y_{(t+1,T)}|S_t^1, \ldots, S_t^m, S_{t-1}^{m+1}, \ldots, S_{t-1}^M, \theta) \end{cases} \tag{1}$$

Thus we have:

$$\gamma_t = P(S_t|Y_{(t)}, \theta) \propto \alpha_t \beta_t \tag{2}$$

α_t and β_t can be computed recursively:

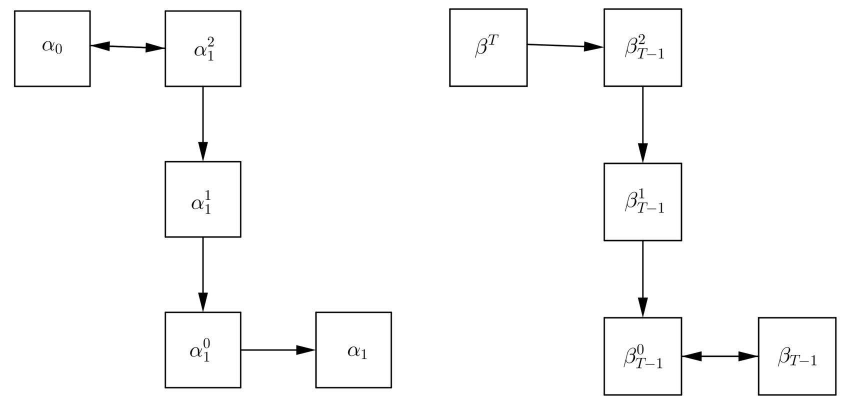

Figure 2: alpha-beta recursions

Gibbs sampling

To go through we do not need to compute the posterior but only $\langle S_t^m \rangle$, $\langle S_t^m S_t^n \rangle$ and $\langle S_{t-1}^m S_t \rangle$. Using the law of large numbers and MCMC methods we can avoid doing exact inference. We use Gibbs sampling here. Since we have a graph structure the iteration gives:

$$\begin{aligned} &P(S_t^m|S_{(-t)}^{(m)}, Y_{(t)}, \theta) \propto \\ &P(S_t^m|S_{t-1}^m)P(S_{t+1}^m|S_t^m)P(Y_t|S_t^{(m)}) \end{aligned} \tag{3}$$

We use an *impatient* Gibbs sampler, that is we only pick few samples and introduce no burn-in period: We do not try to reach convergence since this algorithm can be very slow.

References

[1] Zoubin Ghahramani and Michael I Jordan. Factorial hidden markov models. *Machine learning*, 29(2-3):245–273, 1997.

[2] CF Jeff Wu. On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95–103, 1983.

Results

We implemented all the four described methods and tested our implementation computing the evolution of the log-likelihood (experiments: $K = 2$, $M = 3$).

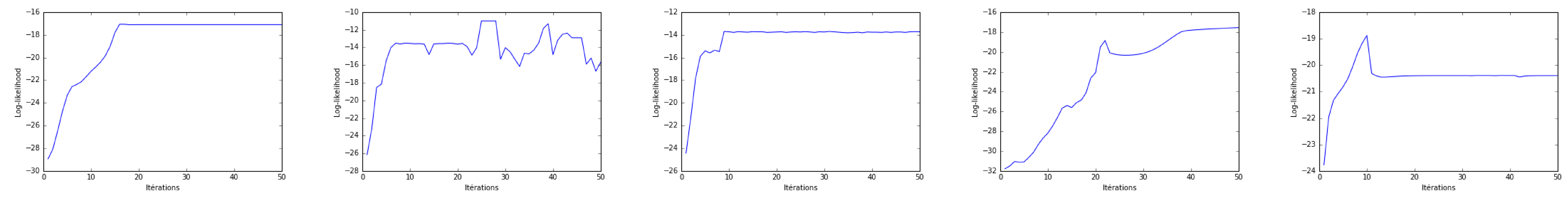

Figure 3: Log-likelihood over EM iterations for the different models (from left to right: exact inference, impatient Gibbs sampling, Gibss sampling with burn-in, mean-field, structured mean-field)

Note that the log-likelihood is not always increasing for the variational models since we compute an approximate log-likelihood.
In [1] the author suggested using only 10 samples and no burn-in for the Gibbs sampling however we found that adding a small burn-in period smoothed our log-likelihood.

E-step method	Iteration duration (s)
HMM- exact inference	0.453
Factorial HMM- exact inference	0.429
Gibbs sampling	1.100
mean-field	0.053
structured mean-field	0.319

Table 1: Computation times

As M increases exact inference became intractable and Gibbs sampling inefficient. Variational methods are still satisfactory though.

Variational methods

Using MCMC we have no information regarding the speed of convergence to the posterior.
We now introduce a new family of methods which put constraints on the E-step thus not yielding the posterior but a *constrained* posterior:

- mean-field: all hidden states are supposed independent,
- structured mean-field: we relax the constraint and suppose we have M independent Markov chains.

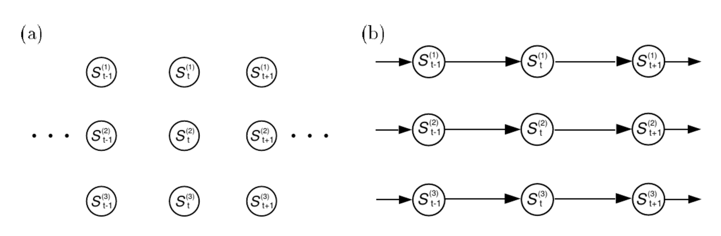

Figure 4: Graphical models for the variational inference (mean-field, structured mean-field)

Conclusion and future work...

In our work:

- We implemented the four methods to perform inference (the E-step) on Factorial HMM,
- we investigated the mathematical model behind Factorial HMM,
- and we experimentally verified the benefits of variationnal models in dealing with large space states.

What's next As next steps, we wish to test our implementation on real data (the Bach Chorales) and compare the properties of this model versus HMM on test data.