Mixed Memory Hidden Markov Models

Çağatay Yıldız, Taylan Cemgil
Boğaziçi University, Computer Engineering Department

Model

Mixed Memory Hidden Markov Model(MMHMM) is a model that extends what is proposed in [1] by assuming an underlying hidden state world generates observations. It also differs from the classical first order HMM in the sense that current hidden state does not depend only upon the last hidden state but one of the K previous ones, where K is a fixed lag.

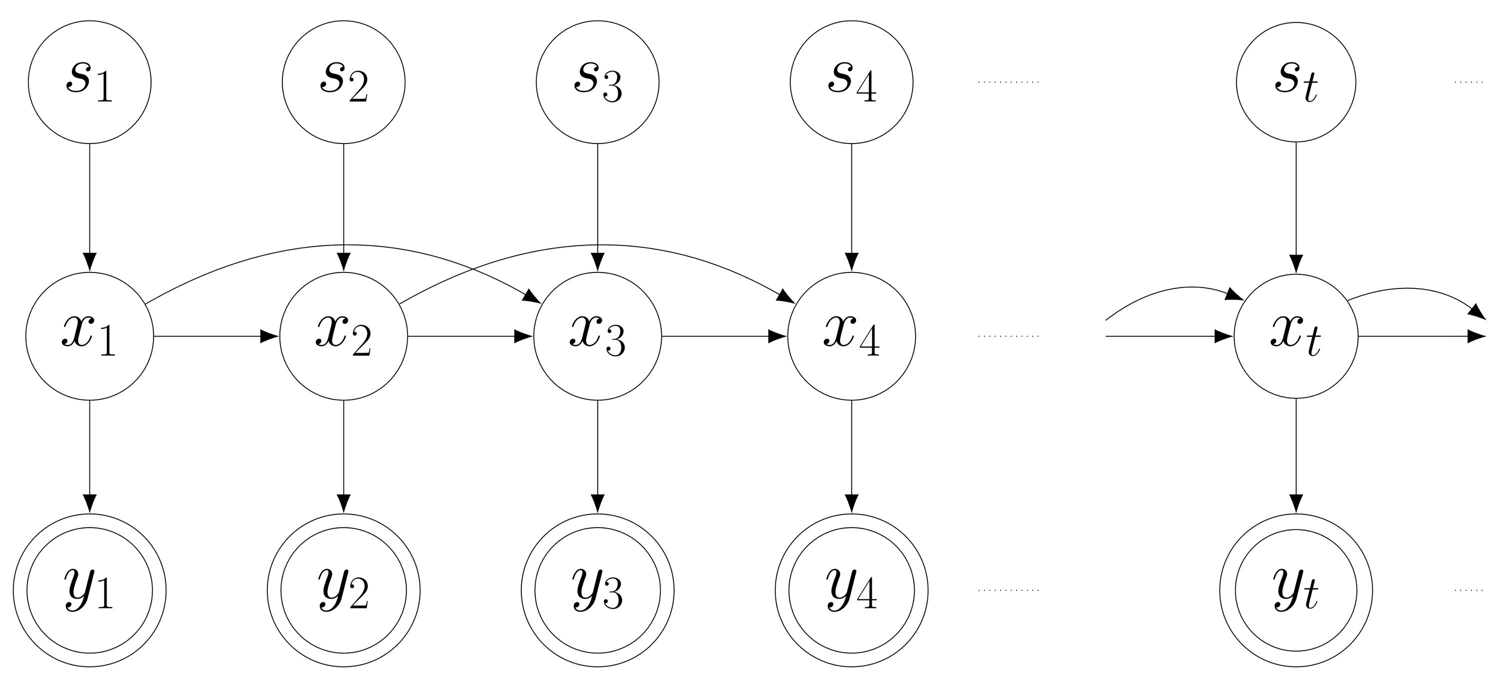
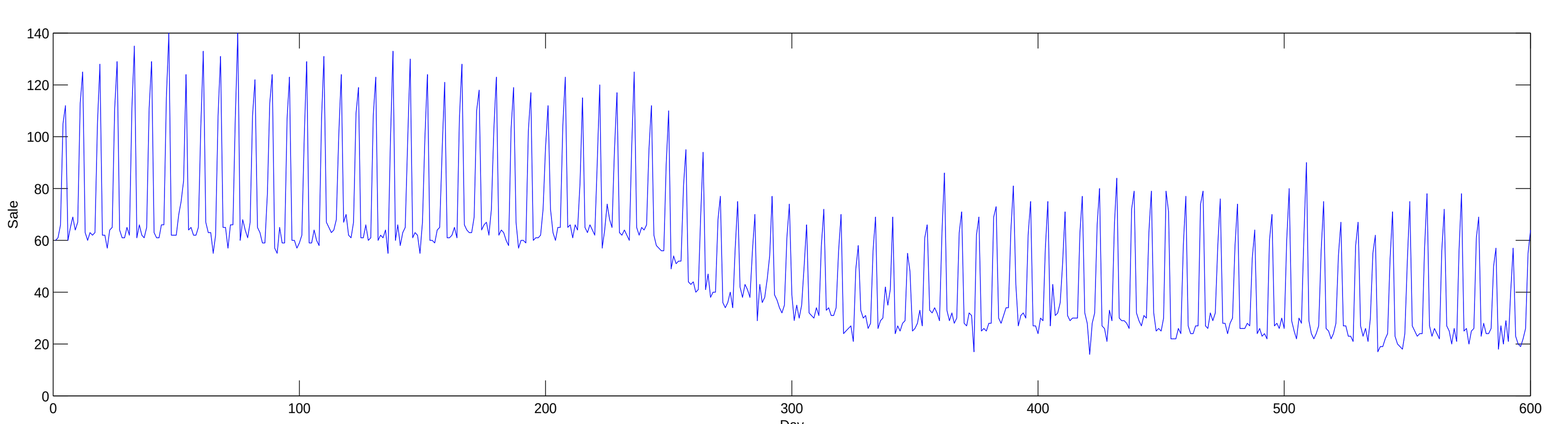


Figure 1: Graphical model for 2nd order mixed memory hidden Markov model. Here, s_t manages to order of transition between hidden states.

Example Application



Above is the illustration of the sale data of a certain product for 600 days. Setting K to an appropriate number, MMHMM can model such regular fluctuations.
Furthermore somewhere around $t = 250$, a certain decrease in sale takes place. If forgetting property is added to MMHMM, this kinds of shifts can also be captured.

EM Updates

$$\pi_i \propto \sum_t p^{old}(s_t = i|Y)$$
$$B_{i,j} \propto \sum_t p^{old}(x_t = i|Y)[y_t = j]$$
$$A_{i,j,w} \propto \sum_t p^{old}(x_{t-k} = i, x_t = j, s_t = w|Y)$$

where

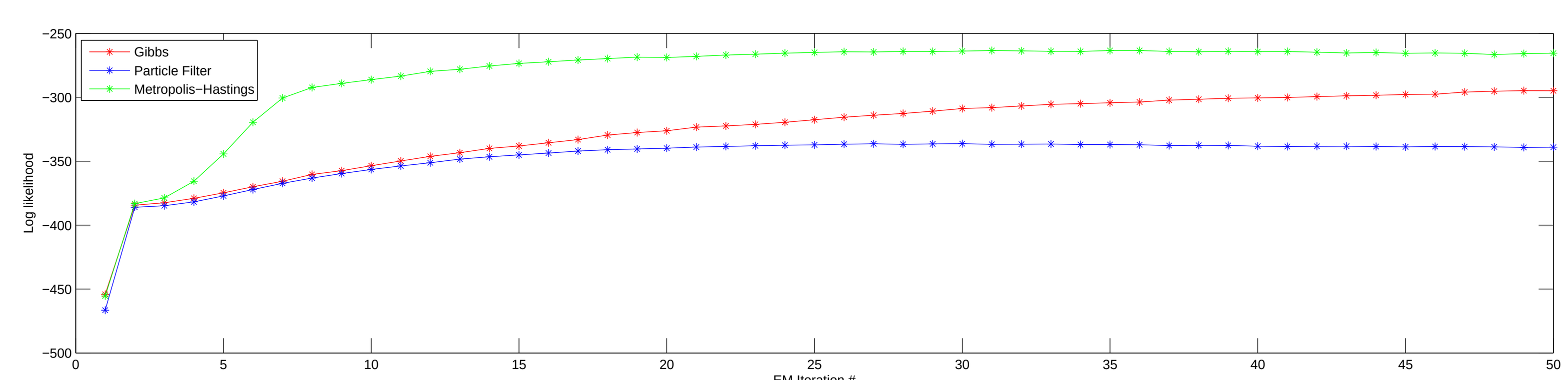
$$p(x_{t-w} = i, x_t = j, s_t = w) = A_{i,j,w}$$
$$p(x_t = i, y_t = j) = B_{i,j}$$
$$p(s_t = i) = \pi_i$$

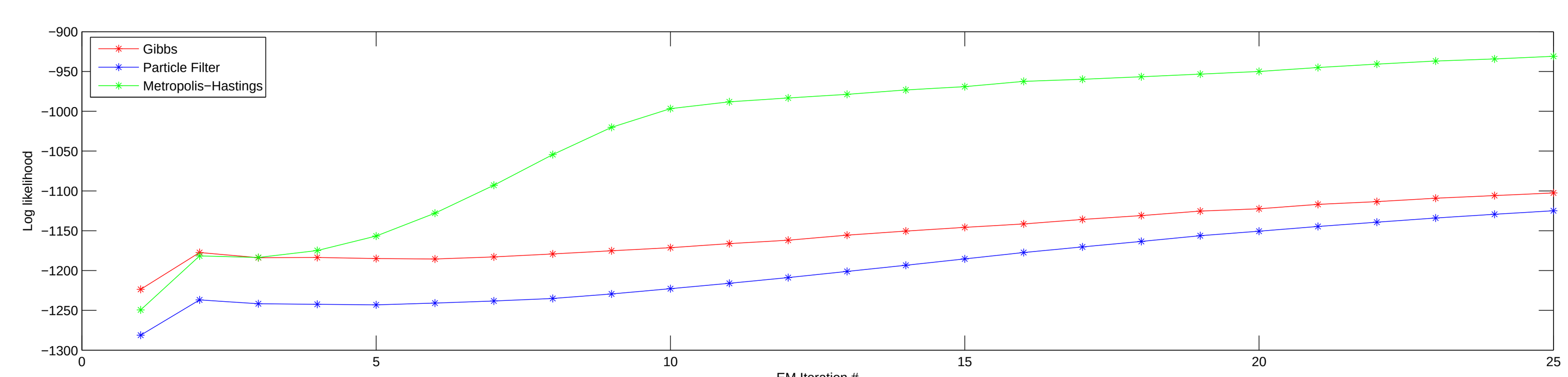
Monte Carlo

To perform EM operations, posterior probabilities are needed. In this work **Gibbs sampler**, **particle filters** and **random walk Metropolis-Hastings** are used for sampling. Because different settings of switches, S, yields different structures to be learned, first $p(S|Y)$ is sampled. Then, depending upon the sampled S values and observations, hidden states, X, are sampled. At the end, the below formulas are used to update model parameters.

$$p(s_t = i|Y) = \approx \frac{1}{N}\sum_{\tau=1}^{N}[s_t^\tau = i]$$
$$p(x_t = i|Y) \approx \frac{1}{N}\sum_{\tau=1}^{N}[x_t^\tau = i]$$
$$p(x_{t-k} = i, x_t = j', s_t = w|Y) \approx \frac{1}{N}\sum_{\tau=1}^{N}[x_{t-k}^\tau = i][x_t^\tau = j'][s_t^\tau = w]$$

Results





Figures show how log-likelihood changes as EM iterates. Above plots are generated using synthetic data and below ones are on real data. For both cases, $K = 2$. Because the real time series is longer, its likelihood and the number of EM iterations are a smaller numbers than the first series.
The main reason why Metropolis-Hastings performed best is that the sample size N is

References

[1] Lawrence K. Saul and Michael I. Jordan.
Mixed memory markov models: decomposing complex stochastic processes as mixtures of simpler ones, 1998.