Parameter estimation of hidden Markov models: comparison of EM and quasi-Newton methods with a new hybrid algorithm.

Authors :
Sidonie Foulon[1,2,3], Thérèse Truong[1], Anne-Louise Leutenegger[2] et Hervé Perdry[3]

Affiliations :
1. Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, 94805, Villejuif, France
2. NeuroDiderot, Inserm, Université Paris Cité, UMR1141, 48 bd Sérurier, 75019, Paris, France
3. Université Paris-Saclay, UVSQ, Inserm, CESP, 94807, Villejuif, France

Keywords :
Hidden Markov models, Baum-Welch, quasi-Newton, L-BFGS-B, SQUAREM, optimisation.

Abstract :
Estimating the parameters of a Hidden Markov Model (HMM) prior to hidden states reconstruction is often a computationally demanding step. A natural method to estimate the parameters of HMM is the Baum-Welch algorithm, a particular case of the Expectation-Maximisation (EM) algorithm. However, other methods exist, particularly the direct maximisation of the likelihood, often employing quasi-Newton methods. The Baum-Welch algorithm always converges to a local maximum, but its convergence is slow. Convergence of "classical" quasi-Newton methods (ex. L-BFGS-B) is generally fast. Still, they can be complicated to implement, and their trajectories can be complex, particularly when the likelihood reaches multiple local maxima. We propose a hybrid algorithm, QNEM, which combines the Baum-Welch algorithm and a quasi-Newton method. The likelihood convexity guides the switch from one method to another.
We compare QNEM with the Baum-Welch algorithm, an accelerated EM algorithm named SQUAREM (Varadhan, 2008, Scand J Statist), and L-BFGS-B. We apply the four algorithms to four examples constructed on different models. Two examples are based on the eruption data of the Old Faithful geyser. The last example is built on a genetic model, where we aim to analyse genome-wide data to identify portions of the genome inherited twice from a common ancestor of an individual's parents. We estimate the parameters of each model with the four algorithms and evaluate their performances.
Our results show that the highest performing algorithm depends on the model, the Baum-Welch and SQUAREM algorithms being sometimes faster than L-BFGS-B and QNEM. The latter has globally good performances, at least as fast as the L-BFGS-B, making it a promising alternative to existing algorithms.