Dreyfus, Neural Networks Methodologies and applications

Learning for Recurrent Networks

The main problem with RNN learning using a descent method comes from the time range of the consequences of changing a weight value. The influence of a weight value on the cost function is not limited to the current time, it propagates. Training an RNN by propagating the computation for each input, computing the weight correction and iterating would be very expensive for long training sequences such as videos, difficult to implement on real time systems.

Two training methods:

* Compute the true gradient with respect to the current weights but change the cost function by truncating the computation period to a sliding window. This is Back propagation through time.
* Approximate the gradient of previous states with respect to the current weights by the values of those gradients with respect to the previous weights == Real time recurrent learning.