Readings in Neuroinformatics

Slaven Cvijetic, 16.11.2018

Learning to Forget: Continual Prediction with LSTM, Cummins Fred, Gers Felix A., Schmidhuber Jürgen, Technical Report IDSIA-01-99, January 1999.

Abstract

Long Short-Term Memory (LSTM) models are a powerful way to circumvent the problem of time lags greater than 5 – 10 discrete time steps between relevant input events and target signals, which is where usual recurrent neural networks fail. The time integration of the backpropagated error depends exponentially on its weights resulting into their vanishing or blowing up. LSTM counters this problem by enforcing a constant error flow through special units, where multiplicative gates learn to open or close access to these units. However, LSTM fail when faced with continual time series with no clear beginning or end. Thus, we developed an extended LSTM model by adding "forget gates" to successfully deal with continual time series. These forget gates learn to reset the memory blocks once their contents are not needed anymore for a LSTM. The forget gate activation factor y^{φ}replaces the standard LSTM constant at the memorizing and forgetting stage. In the Continuous Embedded Reber Grammar and Continual Noisy Temporal Order task, we show that extended LSTM performed better than normal LSTM. Therefore, extended LSTM models prove to be lucrative in all sequential processing tasks, where a hierarchical decomposition might be present, but unknown *a priori*.