**By: Vivek Thakur (15mi424), Komal Mehra (15mi413)**

There are two widely known issues with properly training recurrent neural networks, the vanishing and the exploding gradient. This paper proposes a gradient norm clipping strategy to deal with exploding gradients and a soft constraint for the vanishing gradients problem. The structure of the proposed network is similar to that of a standard multilayer perceptron, with the distinction that we allow connections among hidden units associated with a time delay. Through these connections the model can retain information about the past, enabling it to discover temporal correlations between events that are far away from each other in the data.

One approach for computing the necessary gradients is backpropagation through time (BPTT), where the recurrent model is represented as a multi-layer one (with an unbounded number of layers) and backpropagation is applied on the unrolled model. The effectiveness of proposed solutions provides some indirect empirical evidence towards the validity of the hypothesis, though further investigations are required. In order to deal with the vanishing gradient problem a regularization term is used that forces the error signal not to vanish as it travels back in time.