Simulations and theory of generalization in recurrent networks

Juan Valle-Lisboa

Universidad de la Repblica, Montevideo, Montevideo, Uruguay

Abstract

Despite the tremendous advances of Artificial Intelligence, a general theory of intelligent systems, connecting the psychological, neuroscientific and computational levels is lacking. Artificial Neural Networks are good starting points to build the theory. We propose to analyze generalization of learning in simple but challenging problems. We have previously proposed to concentrate on learning sameness, as we have shown that this is difficult for a SRN. Here we present the results of trying to use a Long-Short Term Memory Network to learn sameness. We show that the LSTM although much more efficient to learn partial examples of sameness fails to generalize to a proportion of the examples. This suggests that LSTM and SRN share a core set of features that make generalization of sameness problematic. By analyzing where the two models fail, we arrive at a proposal of what makes sameness hard to learn and generalize in recurrent neural networks.