Pathological Tasks

Giulio Galvan

8 settembre 2015

1 Introduction

The problems discussed below were first introduced by Hochreiter [1] as difficult tasks for RNNs because they require learning long term correlations and have been used as benchmarks since. A few additions are taken from Martens [2].

2 The noisy memorization problem

The task is the binary classification of input sequences of length p which are of the form: "1α" and "0α", where α is a string of $p-1$ number $\in [2, p-1]$. Number i is fed to the network as the p dimensional vector whose i-th component is 1 and all the others are 0. The difficulty of the problem is to store valuable information for long time intervals (p steps in this case) while discarding irrelevant inputs.

3 The addition problem

The problem consists in performing an addition between two real numbers x_i and x_j in $[-1, 1]$ belonging to a sequence of randomly generated numbers. The difficulty in this problem is that such numbers can be arbitrarily distant in the input sequence, so the learning net must exhibit a long term memory. More specifically the input is a sequence of pairs; each pair is composed of a real number and a marker $\in \{1, 0\}$. The marker is used to select the two numbers in the sequence to add. The prediction is the last value in the output sequence, the target is $\frac{x_i+x_j}{2}$. The prediction y is considered correct if $|y - \frac{x_i+x_j}{2}| < 0.04$.

Sequences have random length, say L, between the minimal sequence length T and $T + \frac{T}{10}$, the position of the first marker is sampled in first $\frac{L}{10}$ positions, the last marker is instead sampled in $[\frac{4L}{10}, \frac{5L}{10}]$

4 The multiplication problem

The problem is very similar to the addition problem, here we select two numbers in the input sequences of real numbers in $[0, 1]$ and we need to predict the product.

5 The XOR problem

Again, the problem is the same as the addition one but the input are binary and we are asked to predict the XOR binary operation. This problem has been found particularly hard for both LSTM and hessian-free methods as reported in [2].

6 The temporal order problem

The input sequences are composed of T randomly chosen symbols in $\{a, b, c, d\}$ except for two randomly selected positions for which the symbols are sampled in $\{x, y\}$. The task is to predict the relative order of the two special symbols, that is $\{xx, xy, yx, yy\}$. A variant of the task is to use three special symbols instead of two. Again, the difficulty of the problem is the possibly distance from the special symbols whose relative order is to be detected.

Riferimenti bibliografici

[1] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997.

[2] James Martens and Ilya Sutskever. Training deep and recurrent networks with hessian-free optimization. In Grégoire Montavon, Genevieve B. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade - Second Edition*, volume 7700 of *Lecture Notes in Computer Science*, pages 479–535. Springer, 2012.