# Title

Feed-forward networks with attention can solve some long-term memory problems

# Author

Colin Raffel, Daniel P. W. Ellis

# Abstract

We propose **a simplified model of attention which is applicable to feed-forward neural networks** and demonstrate that the resulting model can solve the synthetic “addition” and “multiplication” long-term memory problems.

# Issue

1. Simple (using feed-forward neural networks)
2. Good performance

# Method

## Attention

**Step 1**. Given a model which produces a hidden state at each time step, attention-based models compute a “context” vector as the weighted mean of the state sequence by

where is the total number of time steps in the input sequence and is a weight computed at each time step for each state .

**Step 2**. Compute a new state sequence , where depends on , and the model’s output at .

**Step 3**. The weightings are then computed by

where is a learned function which can be thought of as computing a scalar importance value for given the value of and the previous state .

**Intuitively contributes to , contributes to , which forms a contribution chain. It’s very natural to regard as how much contributes to or , then we normalize this importance value and get the weights for .**

## Feed-Forward Attention

In this paper, the model is simplified as follows,

As before, is a learnable function, but it now only depends on . In this formulation, attention can be seen as producing a fixed-length embedding of the input sequence by computing an adaptive weighted average of the state sequence .