

Basics about RNNs

Michael Wollowski

Sliding Window Approach of FF nets

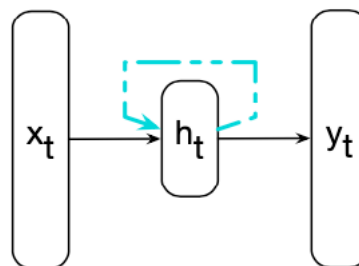
- Feedforward networks use a sliding window of tokens as input.
- We will revisit the sliding window approach when looking at transformers.

Single Input Approach

- RNNs process one word at a time.
- Prior context is stored in recurrent connections,
- This enables a model to potentially depend on information from hundreds of words in the past.

RNNs

- A recurrent neural network (RNN) is any network that contains a cycle within its network connections.
- This means that the value of some unit is directly, or indirectly, dependent on its own earlier outputs as an input.



Clarification on RNNs

Where does the output from the hidden layer go?

- All of the hidden layer units?
- Or just their own unit?

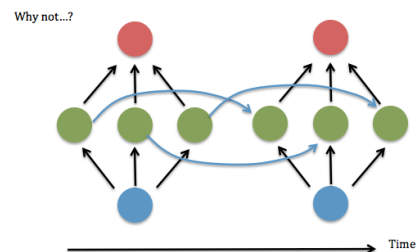
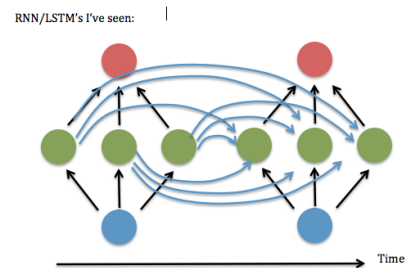


Image source: <https://stats.stackexchange.com/questions/210111/recurrent-neural-network-rnn-topology-why-always-fully-connected>

Clarification on RNNs

- The author of the post conducted three experiments, using working RNN code.
- Experiment 1: set all the non-diagonal terms to zero
it performed terribly
- Experiment 2: fully-connected
it worked well.
- Experiment 3: keeping terms *near* the diagonal, i.e. a banded linear system 3 elements wide
it seemed to work as well as the fully-connected version.

Image source: <https://stats.stackexchange.com/questions/210111/recurrent-neural-network-rnn-topology-why-always-fully-connected>

Clarification on RNNs

- NVIDIA released [this paper, arXiv:1905.12340: "Rethinking Full Connectivity in Recurrent Neural Networks"](#)
- It showed that sparser connections are usually just as accurate and much faster than fully-connected networks.