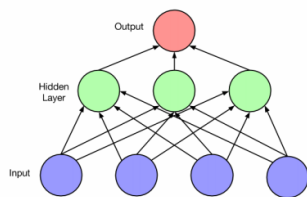


Redes Neurais Recorrentes (RNN e LSTM)

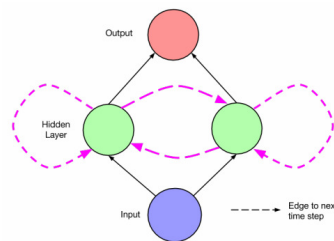
Definição e Exemplos
Prof. Alceu Britto

1

MLP .vs. RNN



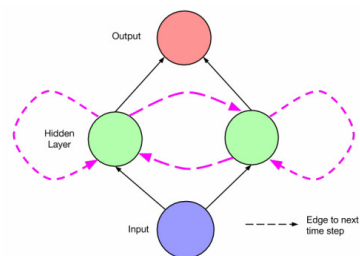
Rede Múltiplas Camadas
(Sem recorrência)
Exemplo MLP – MultiLayer
Perceptron)



Rede Neural Recorrente
(recorrências – linhas
pontilhadas ocorrem ao
longo do tempo)

2

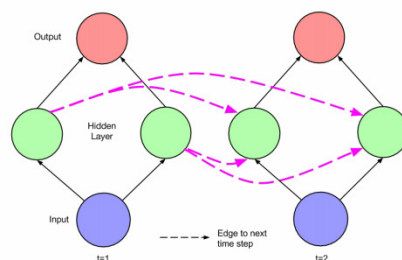
RNN



RNN – visão geral

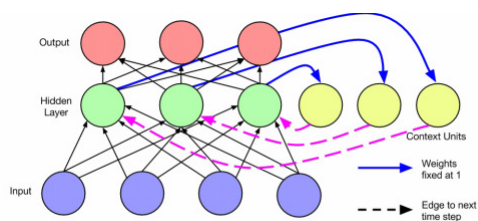
Obs: linhas pontilhadas são recorrências, informação vinda do tempo anterior.

Mesma RNN desdobrada, tempos $t=1$ e $t=2$



3

RNNs precursoras

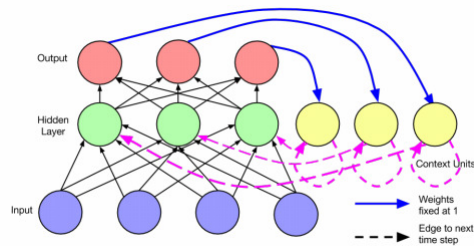


Modelo de Elman de 1990 apresenta *context units*.

Jeffrey L. Elman. Finding structure in time. Cognitive science, 14(2):179–211, 1990.

4

RNNs precursoras



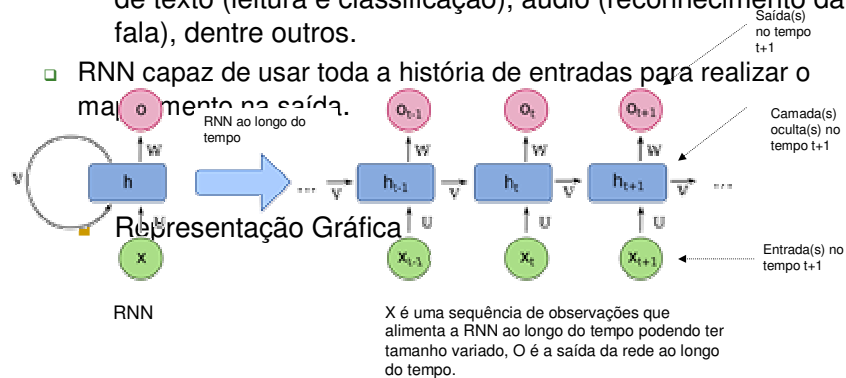
Modelo de Jordan (1996) apresenta *context units* com *autoloop*

Michael I. Jordan. Serial order: A parallel distributed processing approach. Technical Report 8604, Institute for Cognitive Science, University of California, San Diego, 1986.

5

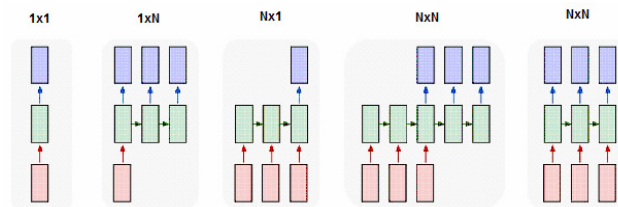
RNN definição

- RNN
 - Aplicadas em sequências temporais
 - Exemplos: previsões de séries temporais, reconhecimento de texto (leitura e classificação), áudio (reconhecimento da fala), dentre outros.
 - RNN capaz de usar toda a história de entradas para realizar o mapeamento na saída.



6

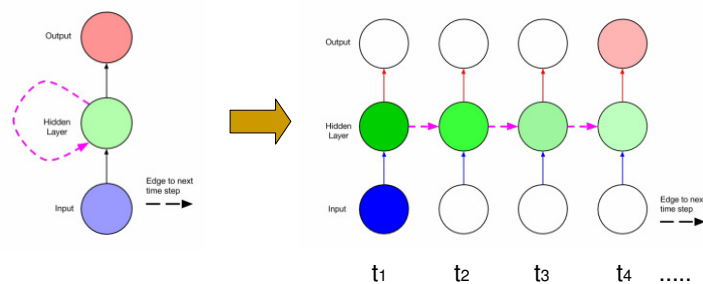
Topologias possíveis



$N \times N$: qtd de entradas x qtd de saídas

7

Problemas: vanishing gradient problem



Contribuição da informação do tempo t_1 decresce exponencialmente ao longo do tempo.

8

Problemas: Memória de Longo Termo em RNNs

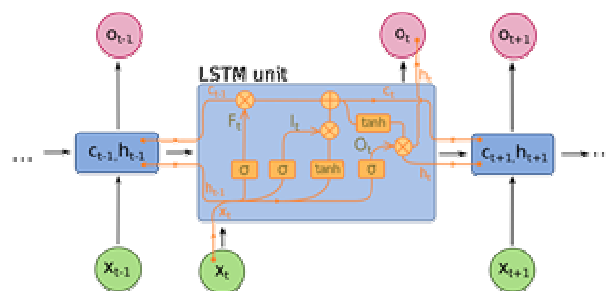
- Suponha que precisamos completar a frase:

As nuvens estão lá no _____.
Eu nasci no Brasil, logo eu falo _____.

- Na primeira frase é mais fácil. Não demanda muita informação de contexto (anterior).
- Na segunda frase precisamos de mais informação contextual vinda de uma memória longa, para prever que devemos completar com a língua materna.
- No link (<http://www-dsi.ing.unifi.it/~paolo/ps/tnn-94-gradient.pdf>) você encontra detalhes sobre a dificuldade das RNNs em tratar dependências de longo termo.
- Qual a alternativa que surgiu?
Uso de RNNs denominadas LSTM (Long Short-Term Memory).

9

LSTM (Long Short-Term Memory)

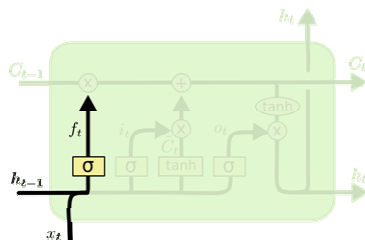


Novidades:
presença de
células (C_t ,
memória); forgetting
gate (F_t); Input gate
(I_t); Output gate (O_t)

Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, 1997.

10

Camada Portão de Esquecimento

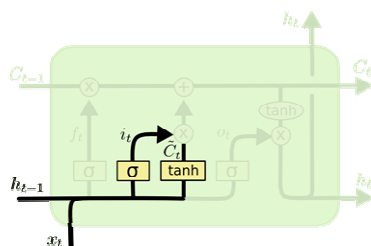


$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

A entrada no tempo t (denominada x_t) e o valor que chega do tempo anterior via camada oculta (h_{t-1}) passa pela camada denominada portão de esquecimento. Uma função sigmoid que emite um valor entre 0 a 1 (sendo 0 esqueça e 1 mantenha tudo). Este valor (f_t) é aplicado no que se encontra na memória (C_{t-1}). Célula é a memória longo termo da LSTM.

11

Adição de Nova Informação na Memória (Célula)



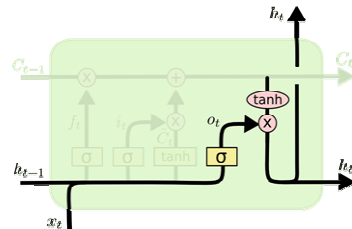
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

A valores de (x_t) e h_{t-1} passam por uma sigmoid e por uma tanh e depois de combinados (produto) são adicionados na memória (C_{t-1}). A célula então é atualizada para C_t .

12

Saída da LSTM



$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$

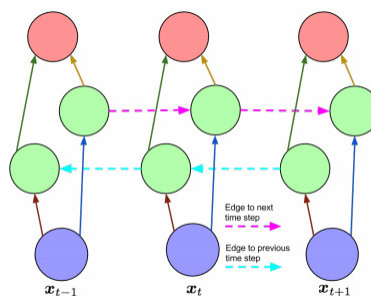
$$h_t = o_t * \tanh(C_t)$$

O valor atual em C_t passa por uma \tanh e o valor da entrada (x_t e h_{t-1}) passam por uma sigmoid. Os resultados são combinados via produto, então temos a saída da LSTM no tempo t (h_t).

13

RNN Bidirecional

- Considera informações do passado e futuro simultaneamente. Muito utilizado no processamento de fala e est



Mike Schuster and Kuldip K. Paliwal. Bidirectional recurrent neural networks. Signal Processing, IEEE Transactions on, 45(11):2673–2681, 1997.

14

Mais Sobre RNN e LSTM

■ Referências

- ❑ Dupond, Samuel (2019). "[*A thorough review on the current advance of neural network structures*](#)". *Annual Reviews in Control*. **14**: 200–230.
- ❑ Sepp Hochreiter; Jürgen Schmidhuber (1997). "[*Long short-term memory*](#)". *Neural Computation*. **9** (8): 1735–1780. doi:10.1162/neco.1997.9.8.1735. PMID 9377276.

■ Links Interessantes:

- ❑ <http://ai.dinfo.unifi.it/paolo/ps/tnn-94-gradient.pdf>
- ❑ <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

15

Exercícios Práticos

■ Exercício 1

- ❑ Previsão de séries temporais (univariada)
- ❑ Previsão de séries temporais (multivariada)
- ❑ Classificação de texto (emoções)

16

Referências

- Lipton, Zachary C.; Berkowitz, John. A Critical Review of Recurrent Neural Networks for Sequence Learning. Disponível em <https://arxiv.org/abs/1506.00019>
- Michael I. Jordan. Serial order: A parallel distributed processing approach. Technical Report 8604, Institute for Cognitive Science, University of California, San Diego, 1986.
- Jeffrey L. Elman. Finding structure in time. Cognitive science, 14(2):179–211, 1990.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, 1997.
- Mike Schuster and Kuldip K. Paliwal. Bidirectional recurrent neural networks. Signal Processing, IEEE Transactions on, 45(11):2673–2681, 1997