

**Data Science  
Academy**

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

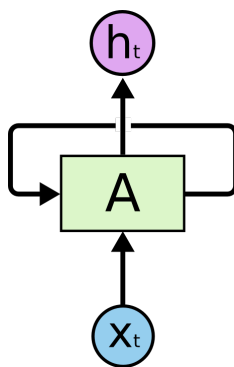
Processamento de Linguagem Natural

O Problema das Dependências  
de Longo Prazo

## Redes Neurais Recorrentes

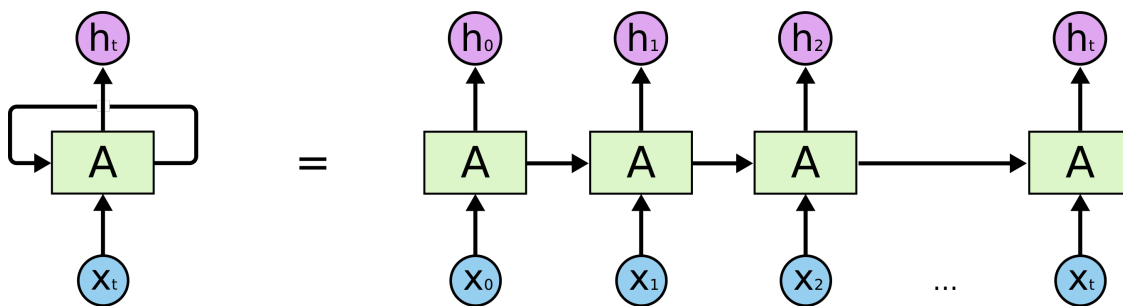
Os seres humanos não começam a pensar do zero a cada segundo. Ao ler este texto, você entende cada palavra com base em sua compreensão de palavras anteriores. Você não lê cada frase e começa a pensar novamente. Seus pensamentos têm persistência.

As redes neurais tradicionais não podem fazer isso, e parece ser uma deficiência importante. Por exemplo, imagine que você deseja classificar o tipo de evento que está acontecendo em cada ponto de um filme. Não está claro como uma rede neural tradicional poderia usar seu raciocínio sobre eventos anteriores no filme para informar os que ocorrem mais adiante. As redes neurais recorrentes abordam esta questão. Eles são redes com loops, permitindo que a informação persista.



No diagrama acima, um pedaço de rede neural, A, analisa alguns dados de entrada  $x_t$  e produz um valor  $h_t$ . Um loop permite que as informações sejam passadas de um passo da rede para a próxima.

Esses loops fazem as redes neurais recorrentes parecerem misteriosas. No entanto, se você pensar um pouco mais, verifica-se que elas não são tão diferentes do que uma rede neural normal. Uma rede neural recorrente pode ser pensada como múltiplas cópias da mesma rede, cada uma passando uma mensagem para um sucessor. Considere o que acontece se observamos o ciclo em mais detalhes:



Essa natureza semelhante a uma cadeia revela que as redes neurais recorrentes estão intimamente relacionadas com sequências e listas. Elas são a arquitetura natural de rede neural a ser usada nesse tipo de dados.

E elas certamente são usadas! Nos últimos anos, houve um sucesso incrível aplicando RNNs para uma variedade de problemas: reconhecimento de fala, modelagem de idiomas, tradução, legenda de imagens e a lista continua. Vou deixar a discussão sobre os feitos surpreendentes que podem ser alcançados com as RNNs para a excelente postagem no blog de Andrej Karpathy, The Unreasonable Efficiency of Recurrent Neural Networks. Vale a pena a leitura:

The Unreasonable Effectiveness of Recurrent Neural Networks

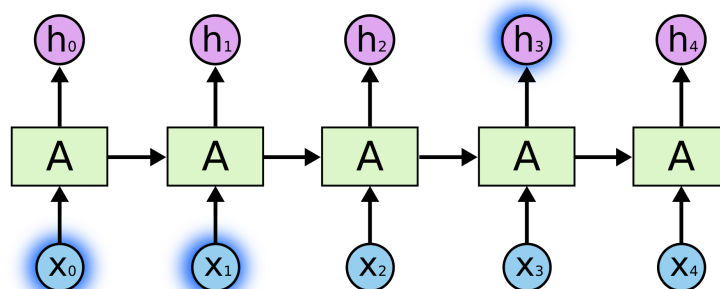
<http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

Essencial para esse sucesso é o uso de "LSTMs", um tipo muito especial de rede neural recorrente que funciona, para muitas tarefas, muito melhor do que a versão padrão. Quase todos os resultados de sucesso baseados em redes neurais recorrentes são alcançados com o uso de LSTMs.

### O Problema das Dependências de Longo Prazo

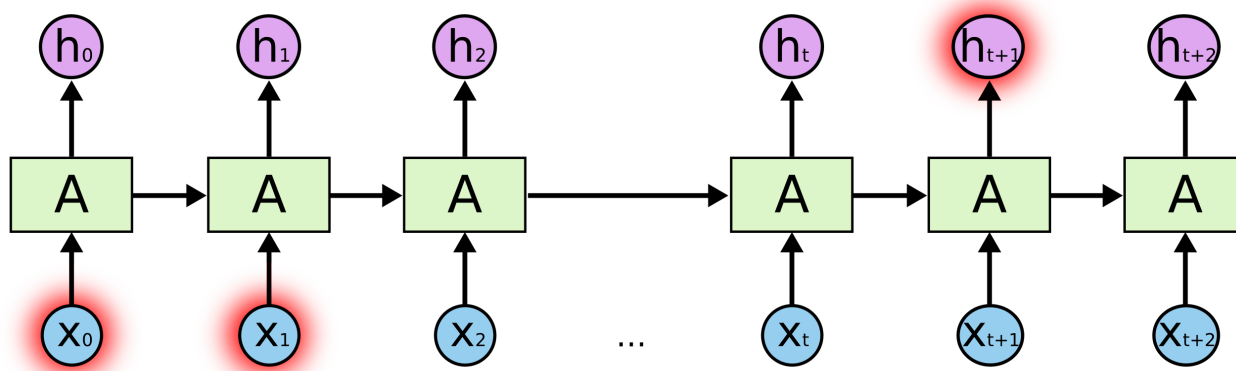
Um dos recursos das RNNs é a ideia de que elas possam conectar informações anteriores à tarefa atual, como usar frames de vídeo anteriores, para compreender o frame atual. Se RNNs pudessem fazer isso, elas seriam extremamente úteis. Mas elas podem? Depende.

Às vezes, precisamos apenas olhar para informações recentes a fim de executar a tarefa atual. Por exemplo, considere um modelo de linguagem tentando prever a próxima palavra baseada nas anteriores. Se estamos tentando prever a última palavra em "as nuvens estão no céu", não precisamos de mais contexto - é bem óbvio que a próxima palavra será o céu. Nesses casos, onde o espaço entre as informações relevantes e o local que é necessário é pequeno, as RNNs podem aprender a usar a informação passada.



Mas também há casos em que precisamos de mais contexto. Considere tentar prever a última palavra no texto "Eu cresci na França ... Eu falo francês com fluência". Informações recentes sugerem que a próxima palavra é provavelmente o nome de uma linguagem, mas se quisermos restringir qual idioma, precisamos do contexto da França, mais adiante. É perfeitamente possível que a diferença entre as informações relevantes e o ponto em que o contexto está, seja muito grande (os "3 pontinhos" na frase do exemplo anterior poderiam ser uma palavra, uma frase ou mesmo um parágrafo inteiro, separando a frase inicial da conclusão do raciocínio e, portanto, definindo o contexto).

Infelizmente, à medida que esse gap aumenta, as RNNs não conseguem aprender a conectar a informação.



Em teoria, as RNNs são absolutamente capazes de lidar com tais "dependências de longo prazo". Um ser humano poderia escolher cuidadosamente parâmetros para que eles solucionassem problemas. Infelizmente, na prática, as RNNs não conseguem aprender. O problema foi explorado em profundidade por Hochreiter (1991) e Bengio, et al. (1994), que encontrou algumas razões bastante fundamentais pelas quais pode ser difícil treinar uma RNN com estes gaps mencionados anteriormente.

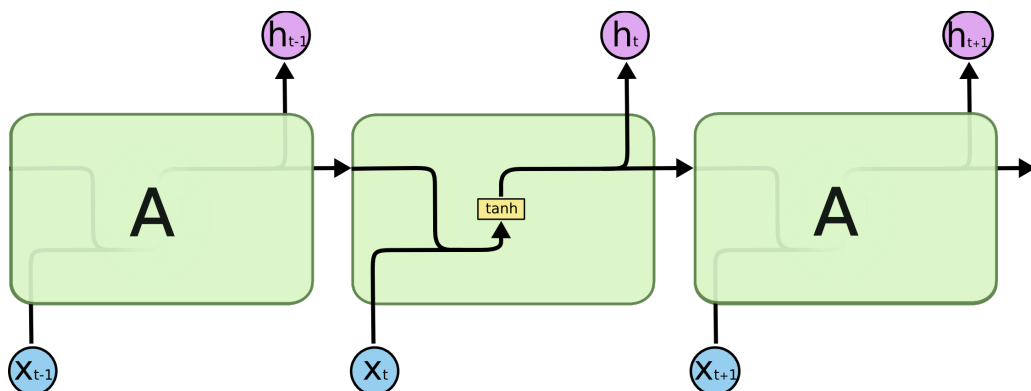
Felizmente, as LSTMs não têm esse problema!

## Redes LSTMs

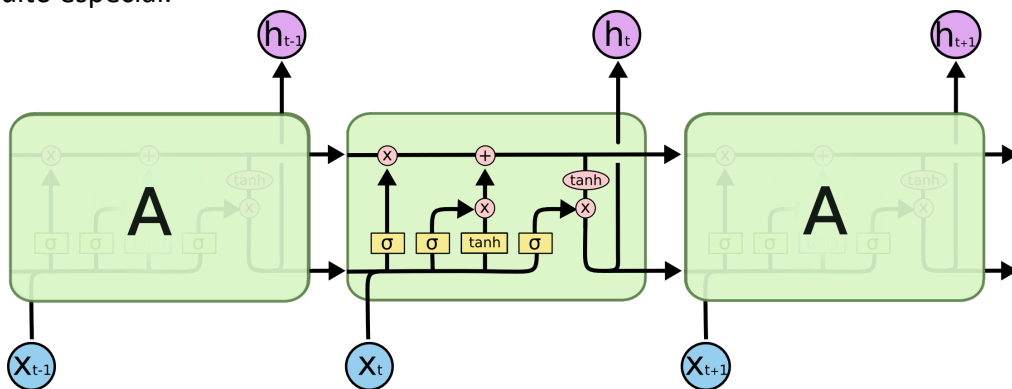
As "LSTMs" (Long Short-Term Memory) são um tipo especial de RNN, capazes de aprender dependências de longo prazo. Elas foram introduzidas por Hochreiter & Schmidhuber (1997), e foram aprimoradas e popularizadas por muitas pessoas ao longo dos últimos anos. Elas funcionam muito bem em uma grande variedade de problemas e agora são amplamente utilizadas.

As LSTMs são projetadas explicitamente para evitar o problema de dependência de longo prazo. Recordar a informação por longos períodos de tempo é praticamente seu comportamento padrão, não é algo que elas lutam para aprender!

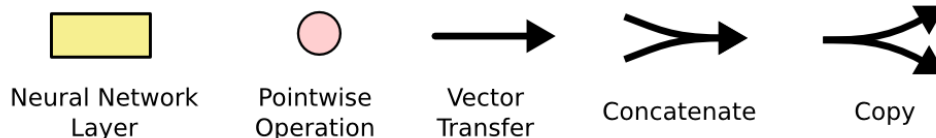
Todas as redes neurais recorrentes têm a forma de uma cadeia de módulos repetitivos de rede neural. Em RNNs padrão, este módulo de repetição terá uma estrutura muito simples, como uma única camada de tanh.



As LSTMs também têm essa estrutura, mas o módulo de repetição tem uma estrutura diferente. Em vez de ter uma única camada de rede neural, existem quatro, interagindo de forma muito especial.



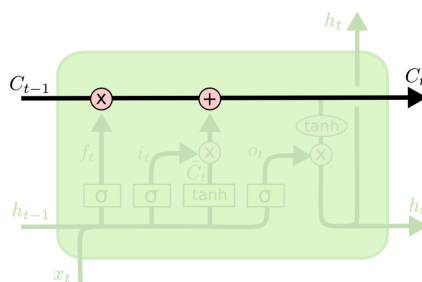
Notação sendo usada:



No diagrama acima, cada linha carrega um vetor inteiro, desde a saída de um nó até as entradas de outros. Os círculos cor-de-rosa representam operações pontuais, como a adição de vetores, enquanto as caixas amarelas são camadas de rede neural. As linhas de fusão indicam concatenação, enquanto uma linha de bifurcação indica seu conteúdo sendo copiado e as cópias indo para diferentes locais.

### A Ideia Central das LSTMs

A chave das LSTMs é o estado da célula, a linha horizontal que passa pela parte superior do diagrama. O estado celular é como uma correia transportadora. Ele corre direto para toda a cadeia, com apenas algumas pequenas interações lineares. É muito fácil que a informação apenas flua ao longo dela inalterada.



A LSTM tem a capacidade de remover ou adicionar informações ao estado celular, cuidadosamente reguladas por estruturas chamadas portões (gates). Os portões são uma forma de permitir opcionalmente a informação. Eles são compostos de uma camada de rede neural sigmóide e uma operação de multiplicação de pontos.

A camada sigmoide produz números entre zero e um, descrevendo o quanto de cada componente deve ser deixado passar. Um valor de zero significa "deixar nada", enquanto um valor de um significa "deixar tudo em ação".

Uma LSTM tem três desses portões, para proteger e controlar o estado da célula.