



**Data Science
Academy**

www.datascienceacademy.com.br

Processamento de Linguagem Natural

Bidirectional LSTMs

Tornar as LSTMs bidirecionais é outra maneira de melhorar a qualidade das previsões de uma LSTM. Com isto queremos dizer: treinar a LSTM com dados lidos do começo ao fim e do fim ao começo. Até agora, durante o treinamento da LSTM, criaríamos um conjunto de dados da seguinte maneira: Considere as duas frases a seguir:

John gave Mary a _____. It barks very loudly.

No entanto, nesta fase, faltam dados numa das frases que gostaríamos que o nosso modelo LSTM preenchesse de forma sensata. Se lêssemos desde o começo até a palavra que faltava, seria como segue:

John gave Mary a _____.

Isso não fornece informações suficientes sobre o contexto da palavra ausente para preencher a palavra corretamente. No entanto, se lermos em ambas as direções, seria o seguinte:

John gave Mary a _____.
_____. It barks very loudly.

Se criarmos dados com essas duas partes, é adequado prever que a palavra que falta deve ser algo como cachorro ou filhote (puppy). Portanto, certos problemas podem se beneficiar significativamente da leitura de dados de ambos os lados. Além disso, isso aumenta a quantidade de dados disponíveis para a rede neural e aumenta seu desempenho. Outra aplicação do BiLSTM é a tradução automática neural, onde traduzimos uma sentença de um idioma de origem para um idioma de destino. Como não há um alinhamento específico entre a tradução de um idioma para outro, conhecer o passado e o futuro do idioma de origem pode ajudar muito a entender melhor o contexto, produzindo assim melhores traduções. Como exemplo, considere uma tarefa de tradução de traduzir filipino para inglês. Em filipino, as sentenças geralmente são escritas com sujeito-objeto-verbo nessa ordem, ao passo que, em inglês, é sujeito-verbo-objeto. Nesta tarefa de tradução, será extremamente útil ler as frases para frente e para trás para fazer uma boa tradução. O BiLSTM é essencialmente duas redes LSTM separadas. Uma rede aprende dados do começo ao fim, e a outra rede aprende dados do final para o começo. Na imagem abaixo, temos a arquitetura de uma rede BiLSTM. O treinamento ocorre em duas fases. Primeiro, a rede de cor azul é treinada com dados criados pela leitura do texto do começo ao fim. Esta rede representa o procedimento normal de treinamento usado para LSTMs padrão. Em segundo lugar, a rede vermelha é treinada com dados gerados pela leitura do texto na direção inversa. Então, na fase de inferência, usamos as informações dos estados azul e vermelho (concatenando os dois estados e criando um vetor) para prever a palavra ausente:

