



**Data Science
Academy**

www.datascienceacademy.com.br

Processamento de Linguagem Natural

Statistical Machine Translation

X

Neural Machine Translation

Machine Translation é a tarefa de converter automaticamente o texto de origem em um idioma para texto em outro idioma.

Em uma tarefa de tradução automática, a entrada já consiste em uma sequência de símbolos em algum idioma, e o programa de computador deve converter isso em uma sequência de símbolos em outro idioma.

Dada uma sequência de texto em um idioma de origem, não há uma única melhor tradução desse texto para outro idioma. Isto é devido à ambiguidade natural e flexibilidade da linguagem humana. Isso dificulta o desafio da tradução automática, talvez uma das mais difíceis na inteligência artificial: o fato é que a tradução requer conhecimento prévio para resolver a ambiguidade e estabelecer o conteúdo da sentença.

Os métodos clássicos de tradução automática geralmente envolvem regras para converter texto no idioma de origem para o idioma de destino. As regras são frequentemente desenvolvidas por linguistas e podem operar no nível léxico, sintático ou semântico. Esse foco nas regras dá o nome a essa área de estudo: Tradução automática baseada em regras ou RBMT.

A RBMT é caracterizada pelo uso explícito e criação manual de regras e representações linguisticamente informadas.

As principais limitações das abordagens clássicas de tradução automática são a especialização necessária para desenvolver as regras e o grande número de regras e exceções requeridas.

Statistical Machine Translation

A tradução automática estatística, ou abreviadamente SMT, é o uso de modelos estatísticos que aprendem a traduzir texto de um idioma de origem para um idioma de destino, dado um grande corpus de exemplos.

Essa tarefa de usar um modelo estatístico pode ser declarada formalmente da seguinte maneira:

Dada uma frase T na língua alvo, buscamos a sentença S da qual o tradutor produziu T . Sabemos que nossa chance de erro é minimizada escolhendo a sentença S que é mais provável, dado T . Assim, desejamos escolher S quanto a maximizar $\Pr(S | T)$.

Esta abordagem oferece uma série de limitações, embora tenha sido a técnica mais avançada até a utilização do NMT. A ideia por trás do NMT é automatizar o processo usado pelo SMT.

Neural Machine Translation

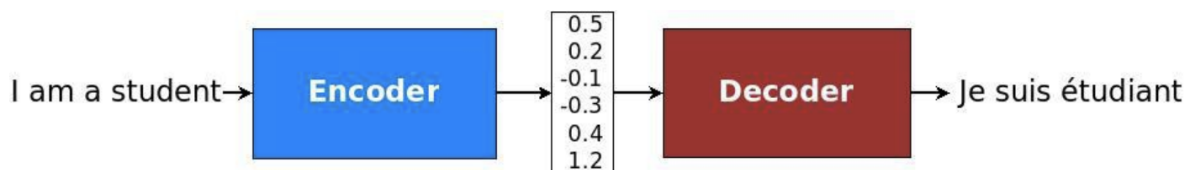
A tradução automática neural, ou NMT, é o uso de modelos de redes neurais para aprender um modelo estatístico para tradução automática.

O principal benefício da abordagem é que um único sistema pode ser treinado diretamente no texto de origem e de destino, não exigindo mais o pipeline de sistemas especializados usados no aprendizado de máquina estatístico.

Ao contrário do tradicional sistema de tradução baseado em frases, que consiste de muitos subcomponentes pequenos que são sintonizados separadamente, a tradução de máquina neural tenta construir e treinar uma única rede neural grande que lê uma sentença e produz uma tradução correta.

Como tal, os sistemas de tradução automática neurais são considerados sistemas ponta-a-ponta, pois apenas um modelo é necessário para a tradução.

A força do NMT reside na sua capacidade de aprender diretamente, de uma ponta a outra, o mapeamento do texto de entrada para o texto de saída associado. O esquema abaixo resume o funcionamento do NMT:



Referências:

Neural Machine Translation by Jointly Learning to Align and Translate

<https://arxiv.org/abs/1409.0473>

Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

<https://arxiv.org/abs/1609.08144>