

Sistema do Google inventou uma língua própria que humanos não entendem

O Google divulgou ontem em seu blog de pesquisa que a inteligência artificial responsável pelo Google Tradutor desenvolveu, sem interação humana, uma língua própria para si. A "descoberta" ocorreu quando a empresa estudava um recurso chamado de "zero-shot translation", e os resultados dessa pesquisa (incluindo a descoberta da língua) foram publicados em um artigo (pdf).

Aula básica de línguas

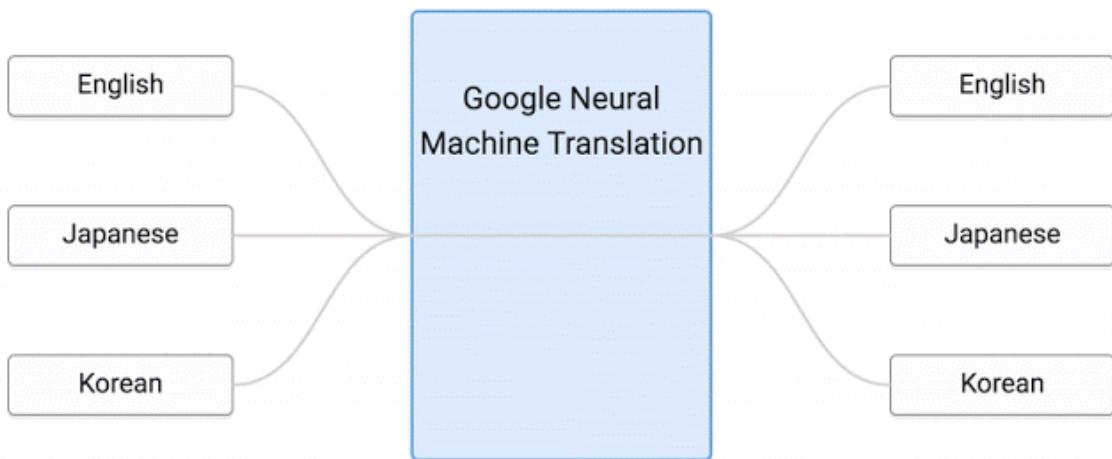
Em setembro, o Google começou a usar redes neurais em seu Tradutor, criando um sistema chamado de Google Neural Machine Translation. Basicamente, o sistema aprende a traduzir a partir de milhões de exemplos diferentes e consegue atingir precisão quase humana. Segundo a empresa, o Tradutor atualmente entende 103 línguas e traduz mais de 140 bilhões de palavras por dia - e aprende com elas.

Para funcionar, o sistema é treinado na tradução de frases específicas de uma língua para outra. Por exemplo, ele precisa traduzir de inglês para coreano, ou de inglês para japonês (e vice-versa). Esse treinamento envolve o uso de milhões de exemplos de traduções previamente validadas, a partir das quais a máquina aprende como a relação de palavras muda entre essas duas línguas.

Aula avançada de línguas

Mas se o sistema é capaz de traduzir do inglês para japonês e do inglês para o coreano, seria ele capaz de traduzir do japonês para o coreano sem passar por inglês? Se sim, essa tradução seria uma "zero-shot translation", e foi ela que motivou a pesquisa dos desenvolvedores do Google Tradutor. O GIF abaixo ilustra essa possibilidade, e a "zero-shot translation" aparece em laranja:

Training

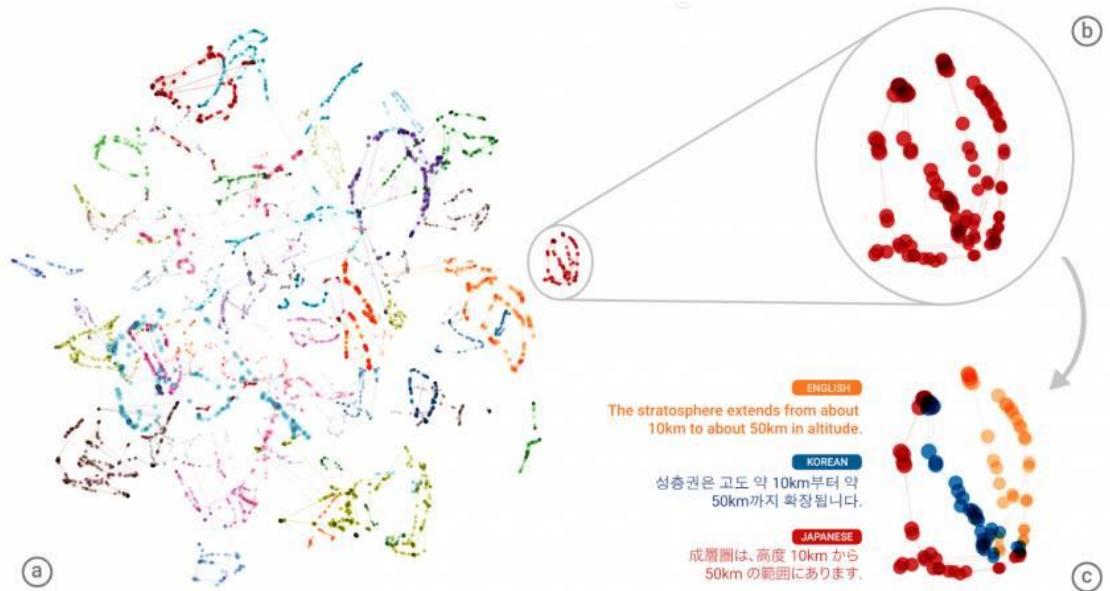


Surpreendentemente, o sistema de tradução conseguiu ir do japonês ao coreano sem ter sido treinado especificamente. "Nós interpretamos isso como um sinal da existência de uma *interlíngua* dentro da rede", disseram os pesquisadores. Em outras palavras, a inteligência artificial teria criado uma língua própria que lhe permite trafegar entre duas línguas nas quais ela não foi previamente treinada, sem precisar passar por uma terceira língua conhecida.

Aula computacional de línguas

De acordo com o *TechCrunch*, essa *interlíngua* parece existir num nível mais profundo de representação. Nesse nível, é possível ver semelhanças entre as línguas diferentes - mesmo as desconhecidas - com facilidade. No entanto, por tratar-se de uma rede neural (sistemas que são normalmente muito pouco acessíveis, até para seus criadores), é difícil dizer mais sobre ela.

Com a finalidade de ilustrar essa "íngua profunda", os pesquisadores criaram uma representação visual das línguas dentro da memória do computador. A imagem abaixo se divide em três partes: **a**, **b** e **c**. A maior delas, **a**, tem pontos coloridos de acordo com significado: uma frase traduzida do inglês para o coreano e uma frase traduzida do japonês para o inglês que tenham o mesmo significado terão também a mesma cor:



À direita, em cima, **b** separa uma dessas frases específicas. Embaixo dela, **c** divide essa frase entre as línguas de origem de cada um dos termos. O que **c** deixa claro é o fato de que, em uma mesma frase, o sistema usa como fonte três línguas diferentes. Isso significa que ele dispõe de um sistema de representação ainda mais fundamental que essas três línguas. Isso, por sua vez, mostra que o sistema está aprendendo algo sobre o significado de cada palavra além de memorizar traduções entre frases.

"Pelo que sabemos, essa é a primeira vez que esse tipo de aprendizagem transferida funcionou em um sistema de tradução de máquinas", disseram os pesquisadores. Para eles, os resultados do estudo são interessantes não apenas para cientistas da computação, mas também para linguístas que pensem em usar máquinas para extrair informações sobre relações entre línguas distantes.

Aula computacional de filosofia

O mais interessante do exemplo usado pelo Google é que a língua inglesa não tem praticamente nenhum grau de parentesco com japonês e coreano. Se houve uma língua comum entre essas duas, ela existiu apenas há muitos milhares de anos. Por isso, com exceção de palavras que são pegadas do inglês pelo japonês ou pelo coreano, não há nenhuma relação etimológica entre as palavras das duas línguas.

Isso significa que a inteligência artificial conseguiu descobrir uma relação entre as palavras de duas línguas não-relacionadas. Se ela foi capaz de fazer isso apenas olhando para milhões de exemplos de traduções entre as línguas, significa que essa relação existe. A chamada *interlíngua* que os

pesquisadores acharam na máquina pode não ser nem mesmo uma língua, mas um sistema ainda mais profundo de representação do que são as nossas línguas.

Link do site: <http://olhardigital.uol.com.br/>

Link da Referência: <http://olhardigital.uol.com.br/pro/noticia/sistema-do-google-inventou-uma-lingua-propria-que-humanos-nao-entendem/64122>