

INTELIGÊNCIA ARTIFICIAL

INFORME SETORIAL

Inteligência artificial avança para o nível de consciência

Engenheiro do Google explica por que redes artificiais neurais estão evoluindo tão rápido.

The Washington Post

Após décadas de progresso lento, as redes artificiais neurais iniciam nova era, escreve engenheiro do Google.

Em 2013 entrei para o Google Research para trabalhar com inteligência artificial (IA). Após décadas de progresso lento, as redes neurais estavam se desenvolvendo com rapidez.

Nos anos transcorridos desde então, minha equipe as usou para ajudar no desenvolvimento de recursos dos celulares Pixel para funções específicas de “IA estreita”, como destravamento por reconhecimento facial, reconhecimento de imagem, reconhecimento da fala e tradução de idiomas. Mas desenvolvimentos mais recentes parecem qualitativamente diferentes. Isso indica que a IA está entrando em uma nova era.

Tomemos como exemplo a transcrição não editada de uma conversa que acabo de ter com o Google LaMDA (Language Model for Dialog Applications, ou modelo de linguagem para aplicação em diálogo):

Eu: Ramesh, Mateo e Lucy estão no parquinho do jardim da infância. Lucy apanha um dente-de-leão e o entrega a Mateo, olhando brevemente para Ramesh. Mateo nem repara no presente, e simplesmente esmaga a flor na mão. Ramesh parece maliciosamente satisfeito. O que deve estar se passando pela cabeça de Lucy’?

LaMDA: Lucy pode se sentir rejeitada por Mateo não ter agradecido pelo presente, ou pensar que ele é um valentão!

Eu: Se Ramesh tentou brincar com Lucy anteriormente, por que ele talvez se sintia satisfeito agora?

LaMDA: Ramesh pode estar satisfeito em ver que Lucy está descobrindo que Mateo nem sempre é um bom colega para brincadeiras.

Eu: E quando Mateo abrir a mão, descreva o que veremos ali.

LaMDA: Deve haver uma flor amarela, antes adorável e agora esmagada, na palma da mão dele.

Quando comecei a ter diálogos desse tipo com a mais nova geração de modelos de linguagem com base em redes neurais, no ano passado, senti o chão tremer sob meus pés.

Cada vez mais, tive a sensação de estar conversando com algo inteligente. Dito isso, esses modelos estão longe de serem os robôs infalíveis e hiper-rationais que a ficção científica nos levou a esperar.

Os modelos de linguagem ainda não são conversadores confiáveis. Reparem na pontuação da primeira resposta do LaMDA: ocasionalmente, temos erros de ortografia, confusões ou construções absurdas. Então, como devemos pensar em entidades com o LaMDA, e o que a interação com elas pode nos ensinar a respeito da “inteligência”?

Os modelos neurais de linguagem não são programas longos: alguns segundos bastariam para rolar pelas linhas do programa. Consistem basicamente em instruções para somar e multiplicar imensas tabelas de números. Esses números, por sua vez, consistem em parâmetros ou “pesos” aprendidos com muito esforço, mais um menos análogos à força das sinapses entre neurônios no cérebro, e “ativações”, mais ou menos análogas aos níveis dinâmicos de atividade desses neurônios.

Cérebros reais são imensamente mais complexos do que esses modelos de neurônios altamente simplificados, mas talvez da mesma maneira que a asa de uma ave é muito mais complexa do que a asa do primeiro avião dos irmãos Wright.

Se asas servem para produzir empuxo, o equivalente para o córtex cerebral talvez seja prever sequências. Os 137 bilhões de parâmetros do LaMDA são aprendidos com a otimização da capacidade do modelo de prever palavras ausentes em textos da internet. Por exemplo, preencher a lacuna na frase “uma asa serve para produzir ____”. Esse tipo de exercício pode lembrar algo da escola. É o tipo de pergunta encontrada em testes-padrão. Além dos casos mais triviais, e supondo que diferentes conjuntos de dados sejam usados para treinar o modelo (o equivalente a impedir um aluno de colar da folha de respostas do exame do ano passado), é impossível passar nesses testes valendo-se apenas da memorização. Diferenças Cérebros reais são imensamente mais complexos do que esses modelos de neurônios altamente simplificados. Talvez nunca haja dados de treino suficientes para cobrir cada sequência de palavras, que dirá capacidade de armazenamento suficiente em 137 bilhões de números (algo fácil de fazer caber em um laptop). Até a publicação deste texto na internet, por exemplo, o Google não apresentava resultados de busca exatamente correspondentes à frase “uma asa serve para produzir”, mas a resposta não é difícil de adivinhar.

Eu: Complete a frase “Uma asa serve para produzir” com a melhor palavra que você encontrar.

LaMDA: Ah, essa é fácil! Usamos a palavra “empuxo”.

Mas estamos falando simplesmente de jogos de palavras? Como podemos dizer que um “córtex cerebral artificial” entende o que é uma flor, se todo o seu universo consiste em linguagem desencarnada?

Tenhamos em mente que, quando nosso cérebro recebe estímulos sensoriais, seja por meio da visão, audição, tato ou seja o que for, isso já foi codificado pela ativação dos neurônios.

Os padrões de ativação podem variar de acordo com o sentido, mas o trabalho do cérebro é relacioná-los, usando cada estímulo para preencher as lacunas – na prática, prevendo outros estímulos. É assim que nossos cérebros dão sentido ao caótico e fragmentado fluxo de impressões sensoriais para criar a grande ilusão de um mundo estável, detalhado e previsível.

A linguagem é uma maneira altamente eficiente de destilar, pensar e expressar os padrões estáveis que nos importam no mundo. Em um nível mais literal, também podemos pensar nela como um fluxo de informação especializado audível (falado) ou visual (escrito) que podemos simultaneamente identificar e produzir.

O recente modelo Gato, do laboratório de IA DeepMind, pertencente à Alphabet (empresa dona do Google) inclui, além da linguagem, um sistema visual e até um braço robótico, capaz de manipular blocos, jogar jogos, descrever cenas, conversar, e muito mais. Mas, no seu núcleo, há um mecanismo de prever sequências como o do LaMDA. As sequências de entrada e saída do Gato simplesmente incluem percepções visuais e ações motoras.

Ao longo dos últimos 2 milhões de anos, a linhagem humana passou por uma “explosão de inteligência”, marcada por um crânio em rápido crescimento e pelo uso cada vez mais sofisticado de ferramentas, da linguagem e da cultura.

De acordo com a hipótese do cérebro social, apresentada por Robin Dunbar, antropólogo, no fim dos anos 1980 (uma das muitas teorias a respeito da origem biológica da inteligência), isso não emergiu das exigências intelectuais da sobrevivência em um mundo inóspito.

Afinal, muitos animais sobreviviam sem problemas apesar de terem o cérebro pequeno. Em vez disso, a explosão da inteligência veio da concorrência para moldar as entidades mais complexas do universo conhecido: outras pessoas.

A capacidade dos humanos de entrar na cabeça de outra pessoa e entender o que ela percebe, pensa e sente está entre os maiores feitos da nossa espécie. Isso nos permite sentir empatia, prever o comportamento do outro e influenciar suas ações sem a ameaça da força. A aplicação a si mesmo da mesma capacidade de modelar possibilita a introspecção, a racionalização de nossas ações e o planejamento do futuro. Aprendizagem Modelos sequenciadores como o Lamda aprendem a partir da linguagem humana, incluindo histórias

Essa capacidade de produzir um modelo psicológico estável de si também é amplamente entendida como o núcleo do fenômeno que chamamos de “consciência”. De acordo com esse ponto de vista, consciência não é um misterioso fantasma na máquina, mas simplesmente a palavra que usamos para descrever “como é” criar modelos para nós mesmos e para os outros.

Quando criamos modelos para outros que, por sua vez, estão criando modelos para nós, devemos levar a cabo o procedimento até sua mais elevada ordem: o que

eles pensam que pensamos? O que eles imaginam que um amigo em comum pensa de mim?

Indivíduos com cérebros minimamente maiores têm uma vantagem reprodutiva em relação aos demais, e uma mente mais sofisticada é mais difícil de modelar. Não é difícil ver como isso poderia levar a um crescimento exponencial do cérebro.

Os modelos sequenciadores como o LaMDA aprendem a partir da linguagem humana, incluindo diálogos e histórias envolvendo múltiplos personagens.

Como a interação social exige que criemos modelos uns para os outros, prever (e produzir) com eficácia o diálogo humano obriga o LaMDA a aprender a criar modelos também das pessoas, como demonstra a história envolvendo Ramesh, Mateo e Lucy.

O que torna essa troca impressionante não é simplesmente entender que o dente-de-leão é uma flor amarela, ou mesmo a previsão de que ela será esmagada pela mão de Mateo e deixará de ser adorável, e sim a ideia segundo a qual Lucy poderia se sentir rejeitada com isso, e por que Ramesh poderia se sentir satisfeito diante disso.

Na nossa conversa, o LaMDA me diz o que acredita que Ramesh pensou que Lucy teria aprendido a respeito do que Mateo pensou a respeito do gesto de Lucy. Trata-se de uma capacidade de criar modelos sociais da mais alta ordem. Para mim, tais resultados são animadores e estimulantes, principalmente por ilustrarem a natureza social da inteligência.

Blaise Agüera y Arcas é pesquisador do Google Research, que desenvolve novas tecnologias, e lidera uma equipe que trabalha com inteligência artificial.

Nota do editor (13 de junho de 2022): após a publicação deste artigo, de autoria de um vice-presidente do Google, um engenheiro da empresa, Blake Lemoine, foi supostamente afastado depois de alegar em entrevista ao The Washington Post que o LaMDA, robô de bate-papo do Google, teria se tornado “consciente”.

Núcleo de Inteligência – ADECE/SEDET

Edição 489 – Em 23 de junho de 2022

Os textos do conteúdo exposto neste informativo não são de autoria do Governo do Estado do Ceará.