

Card_17: Como estamos ensinando os computadores a compreender fotos

Davi Bezerra Barros

Motivações

O vídeo inicia com a pesquisadora Fei-Fei Li demonstrando a capacidade de uma criança de identificar elementos presentes em uma imagem, e a comparando com os sistemas computacionais modernos que carecem desta habilidade em um nível satisfatório, pois apenas enxergam imagens como conjuntos de dados numéricos, e portanto, não podem utilizá-las em contextos complexos como em veículos autônomos, monitoramento de mudanças climáticas, etc. Fei-Fei afirma que a sociedade ainda é **cega**, pois as nossas máquinas mais inteligentes ainda são cegas.

Os desafios da visão computacional

Por 15 anos Fei-Fei desenvolveu um trabalho no campo de machine learning e visão computacional, com o objetivo de ensinar máquinas a nomear objetos, identificar pessoas e inferir geometria das coisas da mesma forma que os seres humanos o fazem, para que elas possam entender as emoções, ações e as relações entre os objetos observados.

Para isso primeiro é preciso ensinar os computadores a ver: tentar identificar seres e objetos pela descrição de sua geometria não é plausível, pois um objeto pode ter infinitas variações de formas e pontos de vista.

Para resolver este problema, Fei-Fei propôs que além de melhorar os algoritmos, deveria também os alimentar com o mesmo tipo de dados visuais de treinamento que uma criança em fase de aprendizado absorve, tanto em quantidade quanto em qualidade. Para isso eles lançaram o projeto ImageNet, que coletou mais de um bilhão de imagens da internet e as rotularam e categorizaram, criando o maior dataset de redes neurais existente até o momento.

Implementação e resultados

O dataset criado através da ImageNet se provou muito compatível com um tipo específico de algoritmo de machine learning; as **Redes neurais convolucionais**. As características das redes neurais convolucionais somadas às arquiteturas modernas de GPUs e CPUs tornaram possível a criação de modelos generativos e preditivos capazes de performar bem em tarefas de reconhecimento de imagens e a interpretação de seus contextos, alcançando com êxito os objetivos da pesquisa.