

APRENDIZADO DE MÁQUINA **ENCONTRA VISÃO COMPUTACIONAL**

ESTAMOS VIVENCIANDO UM MOMENTO HISTÓRICO EM FUNÇÃO DO GRANDE BOOM DA INTELIGÊNCIA ARTIFICIAL JUNTO À VISÃO COMPUTACIONAL. INÚMERAS ÁREAS VÊM SE BENEFICIANDO E REVOLUCIONANDO SEUS MODELOS DE NEGÓCIO: ROBÓTICA, MEDICINA, VEÍCULOS AUTÔNOMOS, DENTRE OUTRAS.

.....
por Luiz M. G. Gonçalves e Esteban Clua
.....

Embora a inteligência artificial e a visão computacional existam há algumas décadas, pode-se afirmar que três fatores permitiram que ambas pudessem passar por uma grande revolução nos últimos anos. Primeiro, há uma grande quantidade de dados sendo produzidos e armazenados (Big Data), ressaltando-se a importância dos desafios lançados, tais como o ImageNet, que deu origem a importantes arquiteturas para a área de visão computacional baseada em deep learning, tais como o LeNet, AlexNet e GoogleLeNet. Em seguida, temos o desenvolvimento de novas e importantes técnicas e arcabouços mostrando-se robustos para a área. Por exemplo, bibliotecas de aprendizado de máquina profundo (deep learning), como Caffe, CNTK, Tensor Flow, Torch e Theano, permitem que seja possível explorar soluções para diferentes contextos de forma rápida e direta. Por último, destacamos a proliferação de Supercomputação a custo baixo, protagonizado por hardware do tipo GPU.

A Visão Computacional (VC), também conhecida como Análise de Imagens, visa extrair informações acerca do mundo tridimensional usando imagens

A Visão Computacional, também conhecida como Análise de Imagens, visa extrair informações acerca do mundo tridimensional usando imagens obtidas por algum processo de captura.

obtidas por algum processo de captura, tais como câmeras ópticas, por exemplo. A VC utiliza técnicas de processamento de imagem (PI) visando realçar para depois abstrair características (features) da cena que sejam úteis para analisar o seu conteúdo tridimensional.

A grande revolução da Inteligência Artificial deve-se ao avanço e desenvolvimento de sistemas com a capacidade de aprender padrões intrínsecos que regem uma determinada tarefa, observados a partir de um grande volume de dados. A capacidade de generalizar os modelos aprendidos, para assim realizar inferências em novos casos,

permite que os modelos aprendidos possam ser sempre incrementados, sem a necessidade de treinamentos novos e customizados. Embora muitos tipos de dados e sinais se beneficiem desse processo, as imagens são consideradas protagonistas desde o início.

A Aprendizagem de Máquina clássica pode ser descrita por duas etapas. A primeira etapa consiste em extrair os dados brutos que servirão de insumo (ex., imagens). Esses dados serão os sinais de entrada, descrevendo características e propriedades que se deseja avaliar no problema. Por exemplo, podem ser usadas as features providas pela VC tradicional. Baseado nas medições de uma amostra que se deseja avaliar, cria-se um vetor descritor que representa a assinatura daquela amostra de acordo com as medições estabelecidas. Feito isso, a segunda etapa consiste em aplicar algum algoritmo de aprendizagem de máquina para a tomada de decisão sobre o conjunto de medições levantadas. A etapa de extração de características é algo não trivial, sendo difícil encontrar formalizações matemáticas que sejam capazes de descrever fenômenos, e até impraticável, dada a grande quantidade de variações de parâmetros que pode ocorrer.

É justamente aqui que as técnicas de deep learning criam um diferencial na solução do problema: substitui-se a etapa de descrição formal e matemática da extração de características tradicional de VC por um processo de treinamento das redes neurais, sendo esse integrado à etapa de tomada de decisão. Este treinamento é capaz de observar as correlações existentes em um grande volume de dados, aprendendo diretamente a partir dos dados brutos quais padrões são mais relevantes, possibilitando assim ajustar os parâmetros das redes neurais. Esta ideia revolucionou as duas áreas através da inserção da operação matemática de convolução, de forma que os parâmetros possam ser aprendidos a partir do treinamento com dados anotados, em vez de serem definidos empiricamente como sempre foi feito em VC até então.

Mas ainda falta muita coisa para fazer. O quão longe se pode chegar, de

onde vêm os dados e como estes são estruturados, o uso de adversarial datasets, são desafios recentemente estudados assim como vários problemas que ainda estão sem solução. Por exemplo, as CNNs modernas podem mudar drasticamente a sua saída quando uma imagem sofre transformações pequenas. Além disso, quanto mais profunda é a rede artificial neural, maiores são suas falhas à generalização. Por fim, o debate ainda continua no momento em que os mais céticos ainda relutam em aceitar que a IA possa encontrar VC de maneira definitiva. ●



LUIZ M. G. GONÇALVES | Possui doutorado em Visão Robótica no Laboratório de Computação Gráfica pelo Programa de Engenharia de Sistemas e Computação da COPPE-UFRJ, Brasil, em 1999. É professor titular do Departamento de Engenharia de Computação e Automação da UFRN, Natal, RN, Brazil, atuando em Visão Computacional e Robótica.



ESTEBAN CLUA | É professor da Universidade Federal Fluminense e coordenador do UFF Medialab. Sua área de atuação está especialmente focada na área de Videogames, Realidade Virtual, GPUs e deep learning. É um dos fundadores do SBGames. Em 2015 foi nomeado NVIDIA CUDA Fellow.