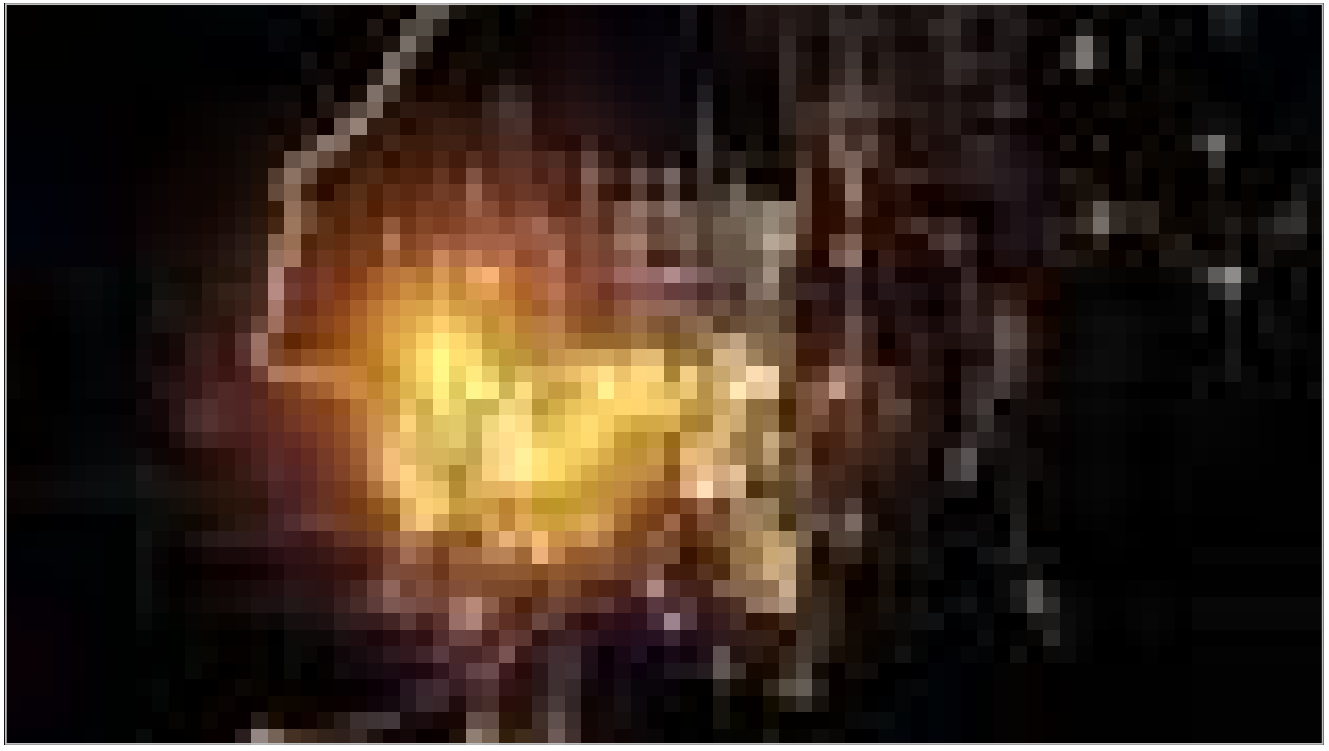
Os riscos do Machine Learning

É importante ressaltar a importância da qualidade das informações apresentadas às inteligências artificiais e quem treina as redes neurais

Wagner Sanchez 12/12/2018 12h15



A inteligência artificial, e consequentemente o *machine learning*, vem sendo objeto de estudo desde a década de 1940, quando surgiram os primeiros neurônios artificiais. Desde então, os desenvolvedores de códigos computacionais vêm aprimorando suas técnicas com o objetivo de simular a inteligência humana e a nossa forma de adquirir conhecimento.

Sabemos que a maioria de nossas habilidades são aprendidas ao longo da infância e da juventude até a vida adulta como por exemplo: falar, andar, escrever, dirigir automóveis, praticar esportes, entre outros. Aprendemos olhando, escutando, interagindo e praticando algo que nos interessa ou que nos é apresentado como necessidade ou como fonte de prazer. Nos algoritmos computacionais, o conceito atual de aprendizado de máquina é semelhante: treinamos as inteligências artificiais sempre que interagimos com uma. As redes neurais artificiais são desenvolvidas para se comportar como uma esponja de absorção de conhecimento, tal qual os bebês quando estão se desenvolvendo, todos sedentos por conhecimentos.

Outra forma de ensinarmos uma máquina é apresentando a ela grandes quantidades de informações em forma de soluções desenvolvidas por seres humanos. Desta forma, as redes neurais artificiais aprendem como resolvemos

problemas e passam a seguir os nossos padrões e, posteriormente, podem adotar seus próprios. Um exemplo interessante deste tipo de aprendizado é a experiência da IBM com o Watson. Ele foi ensinado a criar trailer de filmes através da comparação entre as produções cinematográficas e seus respectivos trailers. Watson “entendeu” como os seres humanos os criam: qual o padrão que os cortes devem seguir de acordo com as expressões faciais dos atores, qual o volume da trilha sonora, quanto tempo deve ter, entre outras características.

Neste contexto, é importante ressaltar a importância da qualidade das informações apresentadas às inteligências artificiais e, principalmente, a idoneidade das pessoas que estão treinando as redes neurais artificiais, pois corre-se o risco de se desenvolver inteligências artificiais preconceituosas, com tendências maléficas. O chatbot Tay, da Microsoft, é um exemplo de como pessoas mal-intencionadas podem ensinar coisas ruins para uma inteligência artificial. Tay foi desenvolvido para interagir e aprender com jovens entre 18 e 24 anos. Porém, alguns delinquentes bombardearam o chatbot com mensagens em prol do nazismo. Com isso, Tay começou tuitar espontaneamente mensagens e imagens a favor de Hitler, como o exemplo a seguir:

Outro exemplo mais recente foi o desafio aceito por um time do laboratório de mídia do MIT - Instituto de Tecnologia de Massachusetts. Três pesquisadores desenvolveram uma inteligência artificial psicopata, deram a ela o nome de Norman, em homenagem a Norman Bates, personagem do filme Psicose de 1960. Para tanto, foram programadas duas redes neurais com o objetivo de interpretar imagens. Para o treinamento destas duas redes foram usadas

imagens com dois padrões diferentes. Para a rede Norman, apresentou-se imagens violentas de cenas de mortes e para a outra rede, foram utilizadas imagens comuns da internet.

O resultado não podia ser diferente. Norman aprendeu a ter comportamento de um psicopata na interpretação de imagens e a outra rede teve uma atuação bem mais amena. Para comprovar o experimento, os pesquisadores utilizaram o teste de Rorschach, usado por psicólogos para a avaliação da saúde mental e emocional dos pacientes. Foram apresentados os borrões de Rorschach às duas redes neurais e o resultado foi surpreendente. Enquanto a rede neural normal interpretava as figuras como sendo pássaros inofensivos, Norman interpretava as mesmas imagens como sendo pessoas mortas de forma violenta. (Este experimento pode ser acessado na íntegra em http://norman-ai.mit.edu/)

Estes dois exemplos ilustram a importância de estarmos atentos ao treinamento das inteligências artificiais que estão presentes em nosso dia a dia: nos atendimentos em hospitais, na área da segurança, na educação, em finanças, em entretenimento etc. Dependendo das informações expostas e das pessoas envolvidas na transmissão de conhecimentos para as redes neurais artificiais, podemos ter surpresas desagradáveis em um futuro muito próximo.