

BOAS PRÁTICAS



Um algoritmo para avaliar a credibilidade de pesquisas

Pentágono quer desenvolver ferramenta capaz de prever se são fidedignos os resultados de estudos em ciências sociais e do comportamento

O departamento de Defesa (DoD) dos Estados Unidos investirá US\$ 7,6 milhões no desenvolvimento de um sistema de inteligência artificial capaz de avaliar a credibilidade de estudos em ciências sociais e do comportamento. Batizado de Score, sigla para Sistematizando a Confiança em Evidências e em Pesquisa Aberta, o programa vai durar três anos. O objetivo do Pentágono é criar uma ferramenta automatizada que atribua uma pontuação a resultados de pesquisas em disciplinas como psicologia, antropologia e sociologia, de acordo com o grau estimado de confiança. A classificação busca orientar usuários de informações científicas sobre o nível de incerteza das conclusões apresentadas.

Segundo o antropólogo Adam Russell, coordenador da iniciativa, o Pentágono com frequência aproveita evidências coletadas por cientistas sociais e psicólogos para desenhar planos relacionados a desafios de segurança nacional,

construir modelos envolvendo sistemas sociais humanos e orientar investimentos. “No entanto, trabalhos empíricos recentes e meta-análises mostram que a capacidade de reproduzir ou replicar esses resultados de forma independente varia de forma dramática”, escreveu Russell, que é gerente de programas da Darpa, a agência de pesquisa do Pentágono. Ele se refere à chamada “crise da reprodutibilidade”, uma sucessão de casos de artigos científicos, sobretudo em áreas como medicina, ciências da vida e psicologia, que caíram em descrédito porque seus resultados não foram confirmados em experimentos subsequentes. Um dos escândalos envolveu Diederick Stapel, professor de psicologia social da Universidade de Tilburg, na Holanda, que teve 30 artigos retratados por manipulação de dados (ver Pesquisa FAPESP nº 190). Há três anos, uma colaboração internacional para investigar estudos de psicologia experimental só conseguiu

reproduzir resultados obtidos em 36 dos 100 artigos avaliados.

No mês passado, a Darpa anunciou que o programa terá como parceiro o Centro para Ciência Aberta (COS), organização não governamental vinculada à Universidade de Virgínia que tem experiência em replicação de experimentos científicos. O COS tornou-se conhecido por liderar, entre 2013 e 2018, a Iniciativa de Reprodutibilidade, um esforço para avaliar se 50 potenciais medicamentos contra câncer descritos em artigos científicos tinham mesmo chance de chegar às farmácias. “Avaliações sobre a credibilidade de pesquisa podem ajudar cientistas a selecionar objetos de investigação, agências a tomar decisões sobre o que financiar e formuladores de políticas públicas a selecionar as melhores evidências”, disse o biólogo Tim Errington, pesquisador do COS.

O programa Score será dividido em quatro fases. Primeiro, será montado um banco de dados com resultados de cerca de 30 mil artigos científicos, abastecido também com evidências extraídas dos próprios *papers* e informações de outras fontes, como o número de citações recebidas por cada trabalho, se seus dados primários estão disponíveis publicamente e se a pesquisa foi pré-registrada – uma garantia de que não houve alteração da hipótese a ser testada durante o curso do experimento. Essa etapa terá a colaboração de pesquisadores das universidades de Siracusa e da Pensilvânia. Em seguida, 3 mil desses artigos serão selecionados e especialistas se debruçarão sobre eles, atribuindo a cada um uma pontuação sobre a possibilidade aparente de terem os resultados replicados.

A ideia é prestar atenção especial em parâmetros que se relacionam com a qualidade dos resultados, como o tamanho de amostras, eventual conflito de interesses, reputação do autor e de sua instituição, entre outros. Depois, os procedimentos usados pelos especialistas para dar pontos a cada artigo serão analisados por cientistas da computação e servirão como base para a construção de algoritmos que reproduzam de forma automática o método adotado.

Por fim, times de pesquisadores tentarão refazer os experimentos dos 3 mil artigos a fim de avaliar se o algoritmo é mesmo capaz de prever se são ou não replicáveis. “A proposta não é substituir homens por máquinas, mas encontrar a melhor maneira de combinar o trabalho dos dois”, disse Russell à revista *Nature*.

Há, naturalmente, o risco de fracasso. A Iniciativa de Reprodutibilidade, liderada pelo COS, mobilizou dezenas de times de cientistas e acabou gerando frutos

restritos. Dificuldades para conseguir reunir condições apropriadas fizeram com que o programa se encerrasse depois de analisar menos da metade dos 50 estudos previstos. Ainda assim, de uma leva inicial de 10 estudos avaliados, apenas cinco foram considerados fidedignos – os demais chegaram a resultados inconclusos ou negativos. Adam Russell, o coordenador do Score, tem experiência em projetos com esse tipo de complexidade. Antes de trabalhar no DoD, gerenciou programas na Iarpa, agência vinculada ao escritório do diretor nacional de inteligência dos Estados Unidos especializada em projetos de alto risco envolvendo inovações. O órgão financia pesquisas em empresas e universidades envolvendo especialistas em matemática, ciência da computação, neurociência, psicologia cognitiva, entre outras.

Para o psicólogo Brian Nosek, professor da Universidade de Virgínia e líder do COS, existe a chance de que o esforço dê errado e o programa falhe em fornecer uma visão fiel sobre a credibilidade de pesquisas, caso não consiga criar um banco de dados robusto ou realizar experimentos de reprodução de alta qualidade. Mas ele considera que o risco vale a pena. “Qualquer que seja o resultado, vamos aprender muito sobre os processos da ciência e como melhorá-los.” ■ Fabrício Marques

O objetivo não é substituir o homem pela máquina, mas combinar o trabalho dos dois, diz o antropólogo Adam Russell



A conexão norte-americana

A controversa experiência de edição de genes de bebês na China, anunciada em novembro em um congresso em Hong Kong, pode ter contado com a participação ativa de um cientista dos Estados Unidos. O biofísico Michael Deem, professor da Universidade Rice, em Houston, está sendo investigado pela instituição por supostamente ter colaborado com o chinês He Jiankui, responsável pelo experimento que gerou dois bebês imunes ao vírus causador da Aids graças ao uso de uma técnica de edição genética, a CRISPR Cas9, em embriões. Pouco antes de fazer o anúncio público, Jiankui havia submetido um manuscrito à revista *Nature* descrevendo o experimento, que foi rejeitado pela publicação. E-mails que vazaram da investigação mostram que o nome de Deem aparecia no final da lista de autores do *paper*, sugerindo que ele foi o supervisor do trabalho. O pesquisador, que nega participação no experimento, já sofre as consequências do escândalo. Sua transferência de Rice para a City University de Hong Kong, onde ocuparia um cargo executivo, foi suspensa.

Deem foi o orientador do doutorado de Jiankui, entre 2007 e 2010. Depois disso, seguiram trabalhando juntos e são coautores de oito artigos publicados. Há indícios de que, em 2017, o norte-americano participou, como membro da equipe liderada pelo chinês, de reuniões com voluntários, ajudando a recrutá-los e obtendo seu consentimento para a pesquisa. Seus advogados argumentam que Deem, embora seja autor de estudos teóricos sobre a técnica, não executou nenhum trabalho experimental, tampouco autorizou a inclusão de seu nome em artigos submetidos a revistas.

Nos Estados Unidos, a geração de embriões geneticamente

modificados é proibida. A eventual participação de Deem em um experimento desse tipo no exterior não seria ilegal, desde que ele tivesse a aprovação de sua universidade, que diz não ter sido informada. Várias nações proíbem o expediente de realizar experimentos eticamente questionáveis em países com regulação frouxa, por meio de colaborações. Recentemente, a União Europeia patrocinou um estudo sobre essa prática, vedada em pesquisas financiadas pelo bloco. O caso mais estranho foi o do neurocirurgião italiano Sergio Canavero, ex-professor da Universidade de Turim, que transferiu para a China sua pesquisa para realizar transplantes de cabeça em seres humanos – que acabou interrompida pelo governo chinês ainda na fase de experimentação animal. Segundo o estudo



ILUSTRAÇÃO NELSON PROVAZI

européu, exemplos semelhantes foram registrados em países da Ásia e da África. “Há casos em que claramente se trata de um artifício para evitar a legislação restritiva dos países da Europa”, disse à revista *The Economist* a professora de ética Doris Schroeder, da Universidade de Central Lancashire, na Inglaterra, que coordenou o levantamento.

Comportamento inadequado de revisor

A revista *Bioinformatics*, da Oxford University Press, publicou um estudo de caso sobre má conduta em revisão de artigos científicos. O curioso é que o episódio envolveu um desvio cometido por um revisor do próprio periódico. Ele sugeriu a autores de um manuscrito a inclusão de mais de 30 citações a artigos dele próprio e deixou claro que o acréscimo influenciaria sua recomendação sobre a aceitação do *paper*. Uma investigação feita pela *Bioinformatics* confirmou o abuso e levou o pesquisador a ser banido do corpo de revisores da publicação.

A revista resolveu reavaliar seus processos para evitar novos problemas e discutiu o tema com o Committee on Publication Ethics,

fórum de editores sobre integridade científica, o que resultou em um novo conjunto de diretrizes. A recomendação para revisores é que se abstenham de sugerir um número exagerado de citações e sejam didáticos e transparentes sobre a importância de incluí-las. Os editores, por sua vez, devem fornecer normas concretas sobre o comportamento esperado dos revisores e verificar se eles estão pedindo a inclusão de seus próprios trabalhos. Já os autores devem avaliar se a inclusão de citações é pertinente e não acatá-la se acharem desnecessária ou imprópria, justificando sua decisão. Caso o revisor reaja de forma negativa, devem informar o editor.