

Enormes bancos de dados sobre compostos e técnicas de garimpagem digital tentam acelerar a descoberta de novas estruturas

Marcos Pivetta

VERSÃO ATUALIZADA EM 03/03/2020

BIG DATA DE MATERIAIS



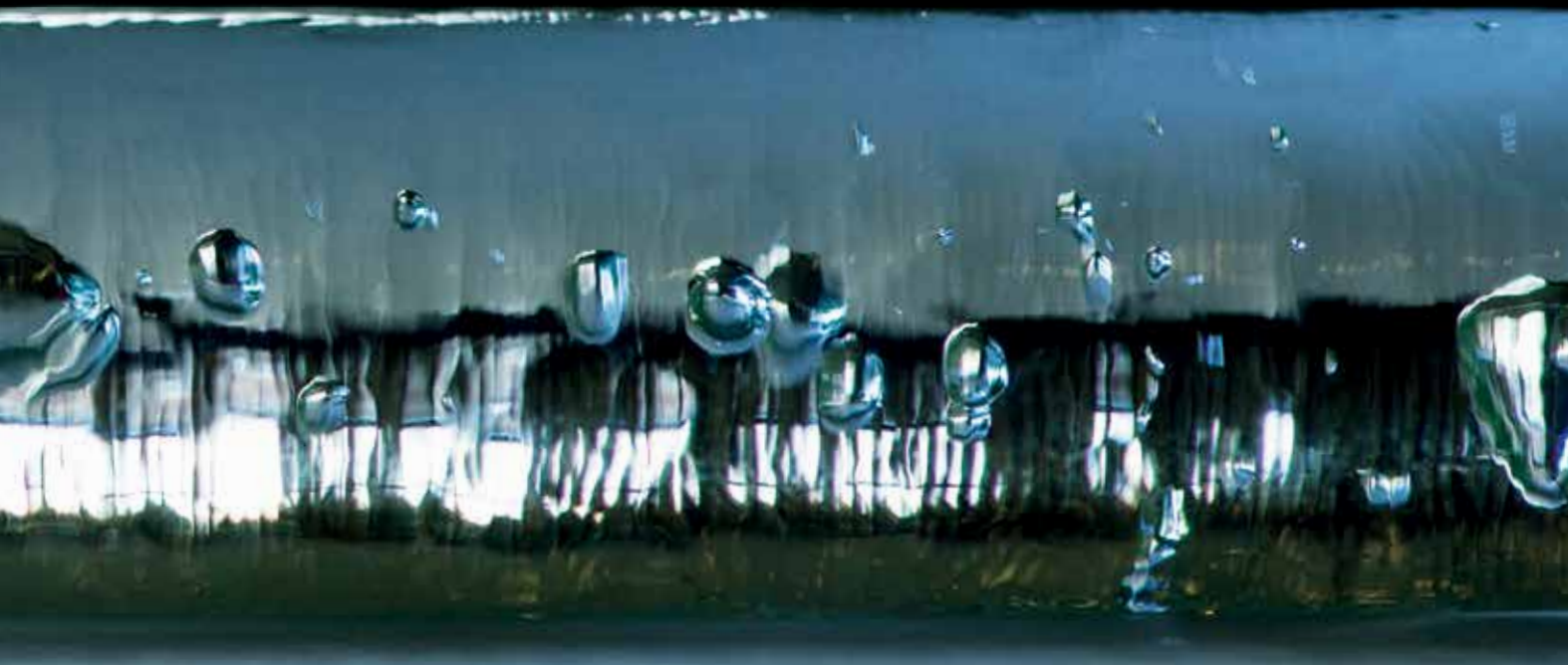
O uso conjugado de abordagens típicas da era do big data pode acelerar a busca por novos materiais e diminuir o caráter empírico desse processo, marcado historicamente por um misto de tentativa e erro, acidentes, observação apurada e, não menos importante, sorte. Pesquisadores de diferentes campos da física, da química e da engenharia de materiais esperam encurtar o caminho que os leve a compostos ainda não encontrados ou sintetizados por meio do emprego de computadores cada vez mais potentes, algoritmos de inteligência artificial e acesso a crescentes bancos de dados sobre propriedades e estruturas de materiais teóricos, nunca feitos em laboratórios, ou reais, já obtidos experimentalmente. Por ora, as promessas são mais estimulantes que as descobertas, mas a abordagem ainda está em sua infância, no exterior e no Brasil.

A equipe do engenheiro de materiais Edgar Dutra Zanotto, da Universidade Federal de São Carlos (UFSCar), contabiliza avanços na busca por novas composições e estruturas vítreas graças ao emprego de ferramentas de inteligência artificial. Em artigo científico publicado no final de janeiro na revista científica *Acta Materialia*, Zanotto e seus colaboradores confrontaram a eficiência de seis algoritmos em correlacionar a composição química de 43.240 vidros óxidos, selecionados em uma base de dados, com uma propriedade fundamental desse tipo de material: a temperatura de transição vítrea (T_g). Esse pa-

râmetro indica a temperatura acima da qual um material amorfo deixa sua fase rígida e quebradiça e passa a exibir um estado mais viscoso e maleável. “Comparamos a eficiência dos algoritmos em prever a T_g de compostos que já conhecíamos e concluímos que dois se destacavam”, explica Zanotto, coordenador do Centro de Pesquisa, Educação e Inovação em Vidros (CeRTEV), um dos Centros de Pesquisa, Inovação e Difusão (Cepid) financiados pela FAPESP. A margem de erro nas projeções fornecidas pelo melhor algoritmo, o Random Forest, foi de no máximo 7,5%, um excelente desempenho, similar ao nível de incerteza das medidas obtidas experimentalmente.

Esse não foi o primeiro trabalho nessa linha da equipe. Há dois anos, Zanotto, o pós-doutorando Daniel Cassar e o professor André Carlos Ponce de Leon, do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-SP), tinham testado a performance de apenas um algoritmo para prever propriedades de vidros. O próximo trabalho vai medir a eficiência dos três melhores algoritmos conhecidos na tarefa de correlacionar a composição química desses vidros óxidos com outras cinco propriedades importantes para aplicações e o desenvolvimento de vidros, como índice de refração e coeficiente de expansão térmica. “Quando os algoritmos se tornarem tão refinados a ponto de antever com precisão as características de um material apenas a partir de sua composição, não precisaremos mais gastar tanto tempo e recursos para testar as

Pesquisadores de São Carlos testam algoritmos que facilitem a busca por novos tipos de vidros



inúmeras possibilidades de descobrir vidros em laboratório”, comenta Zanotto. As formulações não promissoras nem irão para a bancada e os pesquisadores poderão concentrar esforços nos compostos com mais chance de dar certo.

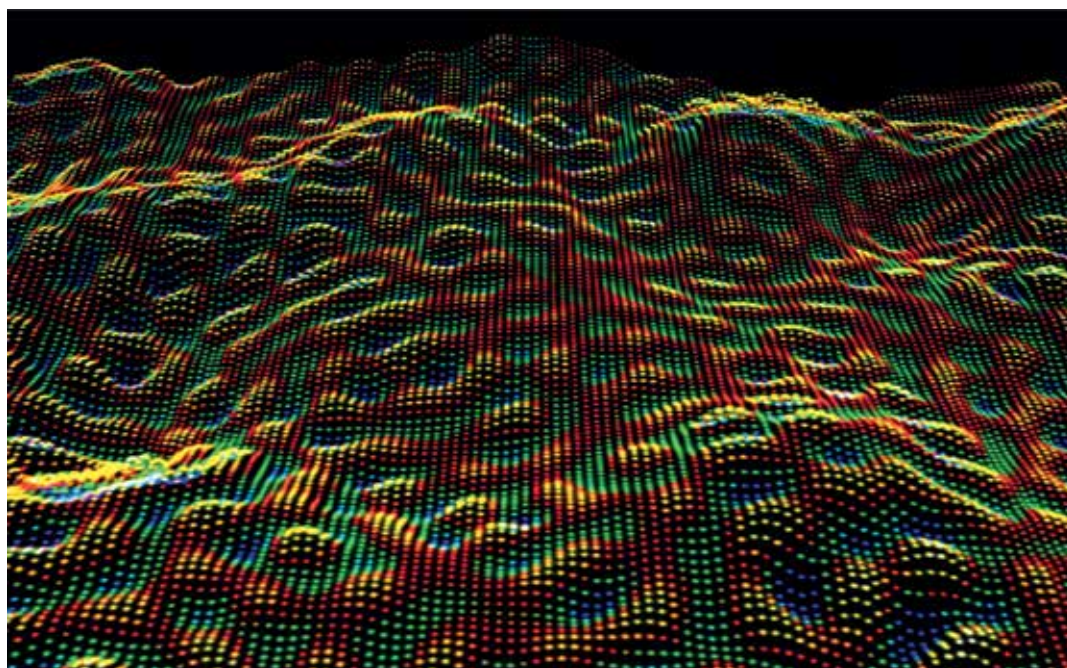
Com a abordagem de big data e as técnicas de inteligência artificial, é possível inverter a forma de trabalhar na ciência de materiais. Historicamente, os pesquisadores primeiro descobrem, por acaso ou depois de incessantes buscas ou modificações, um composto ou uma estrutura inédita e, em seguida, tentam medir suas propriedades e ver se pode servir para algo. Agora o acesso a grandes bancos de dados de materiais permite selecionar as propriedades desejadas e averiguar que tipos de compostos apresentam essas características (ver quadro na página ao lado). Foi isso que fizeram pesquisadores da Universidade Federal do ABC (UFABC) e do Laboratório Nacional de Nanotecnologia (LNNano), do Centro Nacional de Pesquisa em Energia e Materiais (CNPEM). Eles consultaram a base de dados Automatic - Flow for Materials Discovery (Aflow) em busca de compostos tridimensionais que apresentavam uma propriedade quântica associada a um tipo de configuração do spin dos elétrons, o chamado efeito Zeeman.

Até agora, esse efeito, que altera os níveis de energia dos átomos, só foi verificado experimentalmente em materiais bidimensionais, formados por uma única camada de átomos, semelhantes ao grafeno, quando submetidos a um campo magnético. “De uma lista de aproximadamente 59 mil compostos da Aflow, encontramos 20 que apresentam o efeito da forma que queríamos, sem a necessidade de um campo magnético”, comenta o

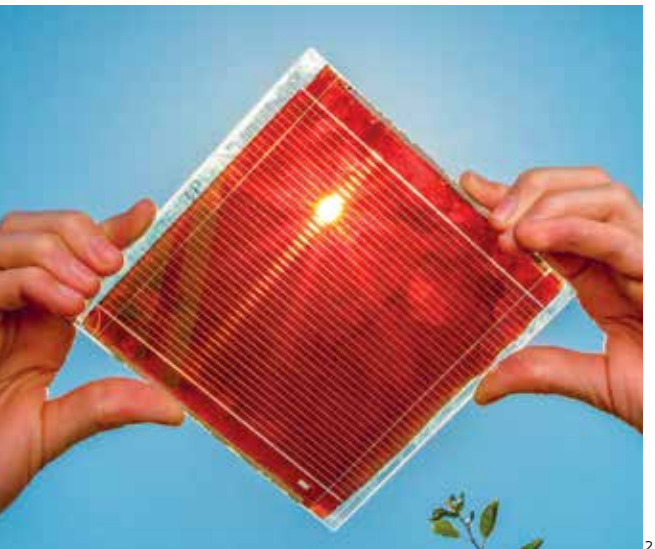
físico Gustavo Dalpian, da UFABC, um dos autores de um artigo publicado no periódico *Quantum Materials* em agosto do ano passado. Essa particularidade confere, teoricamente, uma vantagem para esses materiais na confecção de dispositivos de spintrônica, a eletrônica baseada nos estados do spin (e não na carga do elétron). “Dentro de um aparelho feito com esses materiais tridimensionais que exibem o efeito Zeeman não seria necessário incluir um ímã adicional para gerar o campo magnético. Isso tornaria menos complexo construir esse dispositivo”, explica Dalpian.

Para o físico Adalberto Fazzio, diretor do LNNano e autor de trabalhos sobre o emprego de mineração de dados na busca por novos materiais, as abordagens computacionais são úteis e importantes, porém devem ser refinadas e encaradas de forma realista. “Os algoritmos ainda precisam ser ensinados a encontrar as expressões matemáticas nos bancos de dados que realmente representam os princípios físicos”, pondera Fazzio. Um dos percalços de usar modelos e ferramentas computacionais é obter resultados irreais, que parecem ser um atalho para novas descobertas, mas podem ser uma trilha sem saída. Às vezes, uma simulação aponta como promissores compostos que não são estáveis ou não podem ser fabricados. Nesse caso, só o conhecimento científico, embutido em algoritmos de mineração cada vez melhores ou compilado pela literatura científica, tira o pesquisador do mau caminho.

A julgar pela origem geográfica dos maiores bancos de dados hoje disponíveis, a corrida por novos materiais é liderada por americanos



Superfície de um isolante topológico vista em um microscópio de tunelamento



Novos compostos para células solares são uma das estruturas mais pesquisadas nos bancos de dados

e europeus. “A China tem uma produção que ainda não é muito visível”, diz o físico Osvaldo Novais de Oliveira Junior, do Instituto de Física de São Carlos (IFSC) da USP. “Eles publicam em chinês ou fazem materiais sobre os quais não publicam nada.” Segundo o pesquisador, as técnicas de inteligência de máquina são boas para classificar – sejam imagens, palavras ou propriedades de materiais –, mas não para interpretar. Ao lado de Dalpian, Oliveira Junior editou uma coletânea de artigos sobre o emprego de técnicas de big data na procura por novos materiais em julho de 2019 para a revista *ACS Applied Materials & Interfaces*.

Se esse campo realmente produzir os resultados esperados, histórias como as do médico escocês Alex Flemming (1881-1955), que, em 1928, descobriu de forma fortuita a penicilina, serão cada vez mais raras. Ao voltar de duas semanas de férias, Flemming, que tinha fama de descuidado, percebeu que um mofo esbranquiçado havia se formado sobre um meio de cultura. O fungo evitava o crescimento de bactérias que haviam contaminado inadvertidamente a placa de petri. Assim nasceu o primeiro antibiótico natural. ■

Projeto

CeRTEV – Centro de Pesquisa, Educação e Inovação em Vidros (nº13/07793-6); Modalidade Programa Centros de Pesquisa, Inovação e Difusão (Cepid); Pesquisador responsável Edgar Dutra Zanotto (UFSCar); Investimento R\$ 34.665,855,27.

Artigos científicos

ALCOBAÇA, E. *et al.* Explainable machine learning algorithms to predict glass transition temperature. *Acta Materialia*. 30 jan. 2020.
ACOSTA, C. M. *et al.* Zeeman-type spin splitting in nonmagnetic three-dimensional compounds. *Quantum Materials*. 7 ago. 2019.

Seis bancos de dados

Iniciativas podem ser de caráter amplo ou especializadas na busca por determinados tipos de materiais

THE MATERIALS PROJECT

Lançado em 2011 pelo Departamento de Energia dos Estados Unidos, o projeto é administrado pelo Laboratório Nacional Lawrence Berkeley, na Califórnia. Sua base de dados tem informações sobre a química, a estrutura e as propriedades de 124 mil compostos inorgânicos e 530 mil materiais nanoporosos. Disponibiliza ferramentas que permitem simular as características dos materiais antes de testá-los no laboratório.

AUTOMATIC-FLOW FOR MATERIALS DISCOVERY (AFLOW)

Consórcio formado por 16 universidades (a maioria dos Estados Unidos e algumas da Europa e Ásia), a iniciativa reúne informações sobre 3,2 milhões de materiais compostos, dos quais calculou mais de meio bilhão de propriedades quânticas, térmicas, estruturais e elásticas, entre outras. O Departamento de Energia norte-americano também apoia o projeto, que possibilita a construção virtual de materiais usando tecnologias de varredura de larga escala (high-throughput).

THE NOVEL MATERIALS DISCOVERY (NOMAD)

Repositório europeu capitaneado pela Sociedade Max Planck, da Alemanha, criado no fim de 2015, que reúne oito centros de pesquisas de materiais e quatro de supercomputação. Mantém ferramentas virtuais para pesquisar e cruzar as propriedades, a estrutura e outros parâmetros de milhões de compostos. Tem uma enciclopédia on-line de materiais, com dados sobre uma fração dos materiais encontrados em seu repositório.

HYBRID3

Banco de dados especializado em cristais semicondutores híbridos, formados a nível molecular ou nanométrico por um composto orgânico e outro inorgânico. É tocado por instituições superiores de ensino e pesquisa dos Estados Unidos, sob a liderança da Universidade Duke. Seu foco principal é a procura por materiais que apresentem a chamada estrutura cristalina das perovskitas, similar à do titanato de cálcio (CaTiO₃), que podem ser uma alternativa barata e eficiente para a fabricação de células solares.

COMPUTATIONAL 2D MATERIALS DATABASE (C2DB)

Reúne informações sobre 4 mil materiais bidimensionais, como sua estrutura, elasticidade, propriedades termodinâmicas, eletrônicas, ópticas e magnéticas. Esses materiais são formados por uma única camada de átomos. Seu mais conhecido representante é o grafeno, cuja famosa estrutura hexagonal apresenta a espessura de um átomo de carbono. O banco de dados é organizado por pesquisadores da Universidade Técnica da Dinamarca.

TOPOMAT

Projeto da Escola Politécnica Federal de Lausane, na Suíça, focado numa das classes mais exóticas de materiais, os isolantes topológicos, que conduzem eletricidade em sua superfície, mas, como seu nome indica, comportam-se como isolantes em seu interior. Teve início em 2012 e seu banco de dados contabiliza informações sobre mais de 13,5 mil materiais. Potencialmente, os isolantes topológicos podem levar à produção de novas formas de dispositivos eletrônicos.