



Avalanche de dados

O avanço da *eScience* impacta o modo tradicional de fazer ciência

Yuri Vasconcelos

Foi-se o tempo em que o problema dos cientistas era a dificuldade em obter dados para dar andamento às suas pesquisas. Em muitas áreas do conhecimento, os recentes avanços na área da tecnologia da informação, com a democratização da informática, a ampliação de redes computacionais e a multiplicação das fontes de informação, tiveram como efeito direto uma produção intensiva de dados. Isso ocorre em campos tão distintos quanto a astronomia, inundada diariamente por milhares de imagens e informações de corpos celestes captados por potentes telescópios, a biologia molecular, beneficiada pelo surgimento de máquinas de sequenciamento genético de alto desempenho, e a ecologia, favorecida por uma série de tecnologias e sensores capazes de documentar com precisão as transformações pelas quais passam os diferentes biomas. Tudo isso leva os pesquisadores a se depararem com um novo problema: como fazer para processar, elaborar e visualizar a avalanche de dados adquiridos pelos mais diversos meios. Para dar uma resposta a esse dilema, um novo ramo da ciência tem ganhado cada vez mais importância, a *eScience*, que recorre a modelos matemáticos e ferramentas computacionais para analisar informações e acelerar a pesquisa em outros domínios do conhecimento.

“A ideia de conectar a prática científica tradicional com o acesso, o uso e o processamento de grandes quantidades de dados vai modificar a maneira como fazemos ciência e aumentar sua potencialidade. A FAPESP está na vanguarda desse processo e, no final do ano passado, lançou o Programa *eScience*”, disse Carlos Henrique de Brito Cruz, diretor científico da Fundação, durante o Microsoft *eScience* Workshop

2014, realizado entre os dias 20 e 22 de outubro no Guarujá, no litoral paulista. O objetivo do programa é organizar ou integrar grupos envolvidos na pesquisa sobre algoritmos, modelagem computacional e infraestrutura de dados com equipes de cientistas que atuam em outras áreas do conhecimento, como biologia, ciências sociais, medicina e humanidades.

DESAFIO MUNDIAL

“Uma das principais barreiras que vamos enfrentar talvez seja a dificuldade de comunicação entre os cientistas das equipes necessárias para fazer ciência dessa maneira, fortemente baseada em dados ou em grandes quantidades de dados. Isso requer uma comunicação muito efetiva entre pesquisadores da área de ciência da computação e cientistas de outros domínios. É um desafio no Brasil e em qualquer lugar”, disse Brito. O diretor científico da FAPESP foi um dos conferencistas da mesa-redonda “A importância estratégica da *eScience*”, que também contou com a presença dos cientistas Jason Rhody, diretor sênior do Office of Digital Humanities da National Endowment for the Humanities, e Chris Mentzel, diretor da

Gordon and Betty Moore Foundation, duas organizações norte-americanas que mantêm programas de apoio à ciência.

“Atualmente, todos os campos da pesquisa são afetados pela escala moderna de produção de dados”, disse Mentzel, destacando a importância dos chamados *data scientists* – ou cientistas de dados –, denominação dada aos profissionais que se debruçam sobre o grande volume de dados gerados pelas pesquisas para, a partir dele, produzir novos conhecimentos. “Eles são pesquisadores que trabalham entre as disciplinas. São fazedores de pontes”, disse. Na Gordon and Betty Moore Foundation, Mentzel está à frente de um programa com orçamento de US\$ 60 milhões voltado a incentivar iniciativas na área de *eScience*. Para Rhody, os cientistas estão encarando uma mudança de paradigma. “Estamos passando de uma cultura de escassez para uma cultura de abundância de dados.”

Cunhado em 1999 pelo diretor do Gabinete de Ciência e Tecnologia do Reino Unido, John Taylor, o termo *eScience* recebe outras denominações, como ciência orientada por dados (*data-driven science*) e computação fortemente orientada



por dados (*data-intensive computing*). Alguns países, como Estados Unidos e Inglaterra, já têm programas apoiados pelo governo voltados ao desenvolvimento dessa nova área científica. Aqui no Brasil vale destacar o Núcleo de Pesquisa em *eScience* da Universidade de São Paulo (USP), institucionalizado em 2012. Coordenado pelo professor Roberto Marcondes Cesar Junior, do Instituto de Matemática e Estatística (IME) e um dos coordenadores adjuntos da área de ciências exatas e engenharias da Diretoria Científica da FAPESP, o núcleo é integrado por 20 pesquisadores.

O Microsoft *eScience* Workshop 2014 foi realizado em paralelo à 10ª Conferência Internacional IEEE em *eScience*, organizada pela Computer Society do Institute of Electrical and Electronics Engineers (IEEE), entidade criada nos Estados Unidos por engenheiros eletrônicos e eletrônicos. Durante o evento, foi promovido um painel com pesquisadores contemplados com bolsas do Instituto Virtual de Pesquisas FAPESP-Microsoft Research, que conectam as aplicações da ciência da computação aos desafios da ciência básica em áreas relacionadas às mudanças climáticas e outras disciplinas associadas ao ambiente. Um dos trabalhos apresentados explora soluções inovadoras para o monitoramento de plantas nos trópicos, combinando pesquisa em ciência da computação e fenologia. Um dos ramos mais antigos da ciência, a fenologia é uma área da ecologia que se debruça sobre os fenômenos cíclicos de plantas, como o aparecimento de folhas, botões, flores e frutos, e sua relação com as condições ambientais.

Conectar as aplicações da ciência da computação à fenologia, que estuda os fenômenos cíclicos das plantas



Coordenado pela professora Leonor Patricia Morellato, do Laboratório de Fenologia do Instituto de Biociências da Universidade Estadual Paulista (Unesp), em Rio Claro (SP), o projeto prevê a combinação de tecnologias para monitorar as alterações sofridas ao longo do tempo por vegetações nativas de cerrado, de floresta atlântica, campos rupestres e até caatinga. A área central da pesquisa é em Itirapina, no interior de São Paulo. “Além da observação direta das plantas no solo, instalamos uma câmera no topo de uma torre a 18 metros do solo para fotografar diariamente a vegetação e uma estação meteorológica. Também vamos contar com um veículo aéreo não tripulado (vante) equipado com um sensor hiperespectral e uma câmera para adicionar uma escala espacial à coleta de dados”, diz a pesquisadora. Com elevada resolução espacial, os sensores hiperespectrais podem fornecer detalhes sobre as características físico-químicas e respostas fisiológicas das plantas imageadas. Para Patricia, a fenologia é uma das melhores ferramentas para entender os efeitos das mudanças climáticas nas plantas. “Isso já está estabelecido em regiões temperadas, onde o gatilho da fenologia é a temperatura ambiental e a duração do dia. Mas pouco sabemos sobre o que ocorre nas vegetações tropicais. Com os dados das câmeras e do sensor hiperespectral, queremos definir quais são os gatilhos da fenologia nos trópicos, ou seja, o que faz com que surjam, em determinados momentos, flores, frutos e folhas nas plantas”, diz ela.

ANALISAR IMAGENS

De acordo com Patricia, sem o auxílio de pesquisadores e recursos da ciência da computação seria impossível realizar a pesquisa. “O volume de dados que vamos coletar é gigantesco. Apenas uma câmera digital registra 60 fotos por dia. Temos 11 câmeras monitorando seis tipos de vegetação e precisamos observar a evolução durante, pelo menos, uma estação de crescimento para depois relacionar com o clima. Depois, é preciso processar e analisar todos as imagens, o que seria impossível fazer com uma planilha eletrônica simples. É preciso apoio para trabalhar com esse *big data*. Para isso uma mestrandia criou um banco de dados especialmente para o projeto e um

Sensores sem fio instalados em florestas geram muitos dados sobre processos naturais

pós-doc trabalhou em um *software* para visualização e organização das imagens.”

A pesquisa da professora da Unesp conta com a colaboração do cientista Ricardo Silva Torres, diretor do Instituto de Computação da Universidade Estadual de Campinas (Unicamp), também contemplado com um projeto no âmbito do acordo FAPESP-Microsoft Research. Ele está à frente de um estudo cujo objetivo é desenvolver novas técnicas analíticas e ferramentas computacionais para o processamento de imagens de sensoriamento remoto a fim de analisar a dinâmica de alguns biomas em escalas regionais e continentais. O foco do trabalho, feito em parceria com a professora Marina Hirota, do Departamento de Física da Universidade de Santa Catarina (UFSC), são os biomas tropicais sul-americanos.

Outro trabalho apresentado no evento do Guarujá é liderado pelo ecólogo da Unicamp Rafael Silva Oliveira, que trabalha em parceria com os pesquisadores Antonio Alfredo Ferreira Loureiro, do Departamento de Ciência da Computação da Universidade Federal de Minas Gerais (UFMG), e Stephen Burgess, da University of Western Australia. “Nosso estudo tem como meta investigar a dinâmica de água e carbono em florestas nebulares, pastagens e na área de transição entre elas”, diz Oliveira. Florestas nebulares são encontradas no alto das mon-

tanhas tropicais. “Queremos entender como processos-chave, como absorção e estoque de carbono, transpiração das árvores e captação de água da neblina pela vegetação, são afetados por mudanças no uso da terra e por variações do clima.”

Os estudos de campo ocorrem em uma região de floresta na serra da Mantiqueira, na região de Campos do Jordão, no interior paulista. Segundo Oliveira, está sendo implementada no lugar uma rede de sensores sem fio para monitorar, em três camadas do ecossistema (atmosfera, vegetação e solo), parâmetros microclimáticos de metabolismo da vegetação e de dinâmica da água no solo. “Esses dados poderão melhorar a predição dos impactos ambientais gerados por mudanças no uso da terra e, ao mesmo tempo, possibilitarão o desenvolvimento de modelos hidrológicos e de circulação biosfera-atmosfera com melhor capacidade preditiva”, explica Oliveira. ■

Projetos

1. Towards an understanding of tipping points within tropical South American biomes (2013/50169-1); **Modalidade** Programa Pesquisa em Parceria para Inovação Tecnológica (Pite) e Acordo FAPESP-Microsoft; **Pesquisador responsável** Ricardo da Silva Torres (Unicamp); **Investimento** R\$ 384.838,38 (FAPESP).
2. Combining new technologies to monitor phenology from leaves to ecosystems (2013/50155-0); **Modalidade** Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais – Programa Pesquisa em Parceria para Inovação Tecnológica (Pite) e Acordo FAPESP-Microsoft; **Pesquisadora responsável** Leonor Patricia Cerdeira Morellato (Unesp); **Investimento** R\$ 1.115.752,48 e US\$ 535.902,72 (FAPESP).
3. Interações entre solo-vegetação-atmosfera em uma paisagem tropical em transformação (2011/52072-0); **Modalidade** Programa Pesquisa em Parceria para Inovação Tecnológica (Pite) e Acordo FAPESP-Microsoft; **Pesquisador responsável** Rafael Silva Oliveira (Unicamp); **Investimento** R\$ 644.800,74 e US\$ 663.429,82 (FAPESP).

