

APERFEIÇOANDO A TOMADA DE DECISÕES: USANDO AGENTES INTELIGENTES PARA COMPARAÇÃO DE COMPRAS

Habilidades de software: navegador Web e software de robôs de compra

Habilidades organizacionais: avaliação e seleção de produtos

11.12 Esse projeto lhe oferece experiência prática no uso de robôs de compra para pesquisar produtos on-line, encontrar informações sobre esses produtos e descobrir os melhores preços e fornecedores. Selecione uma câmera digital que você desejaria comprar, como a Canon PowerShot S110 ou a Olympus TG-2 iHS. Visite MySimon (<www.mysimon.com>), BizRate.com (<www.bizrate.com>) e Google Product Search para obter comparações de preços. Compare esses sites de compra em termos de facilidade de uso, número de ofertas, velocidade na obtenção de informações, completude das informações oferecidas sobre o produto e seu fornecedor, e seleção de preços. Qual dos sites você usaria? Por quê? Que câmera compraria? Por quê? De que maneira os sites colaboraram na sua tomada de decisão?

RESOLVENDO PROBLEMAS ORGANIZACIONAIS

As suas redações devem ser avaliadas por um computador?

Você gostaria que os seus trabalhos universitários fossem avaliados por um computador? Bem, pode ser que isso aconteça em seu próximo curso. Em abril de 2013, a EdX, uma parceria entre a Universidade de Harvard e o MIT para desenvolver cursos on-line abertos e massivos (MOOCs, do inglês *Massively Open Online Courses*), lançou um programa de pontuação para respostas dissertativas. Utilizando a tecnologia de inteligência artificial, as respostas dissertativas e as respostas curtas podem ser imediatamente avaliadas e receber um feedback, permitindo que os alunos revisem, reenviem e melhorem a sua nota quantas vezes for necessário. A organização sem fins lucrativos está oferecendo o software livre a qualquer instituição que queira usá-lo. Do ponto de vista pedagógico — se a orientação for bem feita —, o feedback imediato e a capacidade de agir diretamente sobre ele constituem um ambiente de aprendizagem ideal. Mas enquanto os defensores da avaliação automatizada das dissertações defendem a superioridade desse método em relação ao tradicional, em que os alunos aguardam dias ou semanas pelo retorno dos seus trabalhos realizados em papel — os quais eles podem ou não ter a oportunidade de revisar —, assim como pelo benefício da economia de tempo de seus instrutores, os críticos duvidam que os seres humanos possam ser substituídos.

Em 2012, Les Perelman, o ex-diretor de redação do MIT, rebateu um artigo divulgando a proficiência do software de pontuação automatizada de dissertações

(AES, do inglês *Automated Essay Scoring*). O diretor da Faculdade de Educação da Universidade de Akron, Mark Shermis e seu coautor, o cientista de dados Ben Hamner, utilizaram os programas AES em nove empresas, incluindo a Pearson e McGraw-Hill, para avaliar mais de 16 mil respostas dissertativas das escolas de ensino fundamental e médio em testes padronizados de seis diferentes estados norte-americanos. Seu estudo, patrocinado pela Hewlett Foundation, descobriu que a avaliação automatizada ficou bem próxima à avaliação humana, e em alguns casos, produziu uma avaliação mais precisa. Perelman, no entanto, descobriu que não foi realizada nenhuma comparação estatística direta entre os avaliadores humanos e os programas. Enquanto Shermis admite que a análise estatística apropriada não foi realizada — porque as empresas de software impuseram essa condição a fim de permitir que ele e Hamner testassem seus produtos — ele previsivelmente acusa Perelman de julgar o seu trabalho sem ter, ele próprio, realizado nenhuma pesquisa.

Perelman tem, na realidade, conduzido estudos sobre o avaliador de ensaio eletrônico (*e-Rater*) desenvolvido pela Educational Testing Service (ETS) — a única organização que lhe permitiria seu acesso. O *e-rater* utiliza a variedade sintática, a estrutura do discurso e a análise de conteúdo, e é baseado na tecnologia de processamento de linguagem natural. Ele aplica análise estatística às características linguísticas, como elaboração de argumento e variedade sintática para

determinar as pontuações, mas também fornece um peso ao vocabulário e ao conteúdo tópico. Durante o mês em que lhe foi concedido acesso, Perelman analisou os algoritmos e brincou com o *e-rater*, confirmando suas críticas anteriores. O grande problema com os programas AES (até agora) é que eles não podem distinguir o fato da ficção. Por exemplo, em resposta a uma questão dissertativa sobre as causas para o aumento acentuado no custo do ensino superior, Perelman escreveu que o principal responsável foram os professores adjuntos gananciosos cujos salários eram seis vezes maiores do que os diretores da faculdade com pacotes de benefícios exorbitantes que incluíam férias nos Mares do Sul, jatos particulares e contratos de cinema. Ele completou o argumento com uma linha da poesia "Uivo" ("Howl"), de Allen Ginsberg, e obteve nota seis. As métricas que incorreram nessa nota incluíram o comprimento total da redação, o comprimento do parágrafo, o número de palavras por frase, o comprimento da palavra e o uso de advérbios de conjunção como "entretanto" e "além disso". Como os programas de computador não conseguem adivinhar o significado, o comprimento da resposta indica a fluência da escrita, o uso do advérbio de conjunção indica o pensamento complexo, e as longas palavras indicam o conhecimento de vocabulário.

Fornecedores de programas, tais como a Pearson e a Vantage Learning, defendem esses parâmetros, alegando que eles estão altamente correlacionados. Bons escritores adquiriram habilidades que lhes permitem escrever mais sob condições de restrições de tempo; eles usam um vocabulário mais complexo; e sabem como introduzir, interromper, conectar e concluir ideias complexas — as funções dos advérbios de conjunção. Os programas AES também reconhecem fragmentos de frases e avaliam negativamente as frases dos alunos que começam com "e" ou "ou". No entanto, escritores profissionais sabem como empregar ambos com grande efeito. Perelman e um grupo recém-formado de educadores, Profissionais Contra a Avaliação Automatizada de Redações de Estudantes em Avaliações de Alto Risco, alertam que a instrução escrita terá seu nível cultural reduzido para satisfazer às métricas limitadas e rígidas que as máquinas conseguem medir.

Os ganhos de produtividade com o uso do software de avaliação automatizada de dissertações sem dúvida eliminarão alguns postos de trabalho dos avaliadores contratados pelas empresas que aplicam exames padronizados. A Pearson, por exemplo, paga ostensivamente para seus avaliadores entre 40 a 60 dólares por hora. No período de uma hora, um avaliador deverá corrigir entre 20 a 30 dissertações — o que significa de dois a três minutos (e dólares) por dissertação. Claramente, os avaliadores devem usar alguns tipos de métricas simples, a fim de avaliá-las rapidamente. Mas, no mínimo,

eles conseguem reconhecer como falsa a afirmação de que no dia 4 de julho de 2013 os Estados Unidos comemoraram seu 2013º aniversário, mesmo que ela esteja inserida em uma frase bem construída. Enquanto o *e-rater* pode avaliar 16 mil dissertações em 20 segundos, ele não consegue reconhecer que a afirmação anterior é falsa. E, presumivelmente, uma dissertação de 716 palavras contendo várias frases sem sentido não obterá nota seis de um avaliador humano, enquanto uma dissertação mais curta, de 150 palavras, fatural e bem fundamentada, obtém nota cinco, como Perelman foi capaz de demonstrar.

O ETS, desenvolvedor dos exames padronizados SAT, GRE, Praxis e K-12 para vários estados norte-americanos, contra-ataca alegando que o *e-rater* não está substituindo os avaliadores humanos em avaliações de alto risco; ele está complementando. As dissertações são avaliadas por ambos, avaliador humano e máquina, e quando as pontuações não correspondem, um segundo avaliador humano quebra o impasse. Além disso, eles postulam que o curso preparatório para o exame que Perelman desenvolveu para ensinar os alunos a vencer o software AES requer habilidades de pensamento de nível superior — precisamente aquelas que os exames procuram mensurar. Assim, se os alunos conseguirem dominar as técnicas de Perelman, muito provavelmente obterão nota seis. A Pearson acrescenta que seu avaliador inteligente de dissertação é basicamente uma ferramenta de sala de aula, permitindo que os alunos revisem suas dissertações várias vezes antes de entregá-las a um professor para serem avaliados. Mas para muitos estados norte-americanos que querem introduzir seções dissertativas em sua bateria de testes padronizados do K-12, e para aqueles que abandonaram o esforço devido ao custo, eliminando completamente os avaliadores, os avaliadores automatizados os tornarão acessíveis. E os riscos não são insubstanciais no caso do fracasso em alcançar as notas de aprovação no exame padronizado do estado norte-americano, que vão desde a aplicação de um novo exame a programas de recuperação, cursos de verão e à não promoção.

A ferramenta livre EdX parece ser mais sofisticada do que algumas ofertas de fornecedores na medida em que é "treinável" com pelo menos alguma capacidade de desenvolver padrões de classificação e adaptar às preferências de classificação. Em primeiro lugar, os instrutores avaliam 100 dissertações ou questões dissertativas, e essas são a entrada para o programa. Utilizando essas orientações, a ferramenta desenvolve métricas de classificação personalizadas e segue o método de pontuação preferido pelo instrutor, seja um sistema numérico ou uma classificação por letras. Como observado por Shermis, em muitas faculdades de classificação inferior à dos críticos, as classes são agora tão numerosas a ponto de tornar inviável o feedback

abrangente de dissertações. Além disso, nas universidades de primeira linha, o nível de instrução é maior com menos alunos carentes de reforço. Bem nas trincheiras de ensino, uma ferramenta que pode simular adequadamente a avaliação humana, sem maior variação do que a observada de instrutor para instrutor, e que fornece orientação imediata, é um acréscimo bem-vindo à caixa de ferramentas de ensino. Mas, à medida que diminuem as demandas do tempo do instrutor, será que os administradores das universidades efetuarão cortes de pessoal para atender restrições orçamentárias? Será que cada vez menos instrutores estarão ensinando cada vez mais alunos?

À medida que o MOOC e a AES proliferam, a resposta é: muito provavelmente, sim. A EdX está se tornando rapidamente controversa nos círculos acadêmicos. Atualmente, os cursos oferecidos através dela são gratuitos e os alunos ganham um certificado de conclusão, mas não ganham crédito de curso. Para tornar-se autossustentável, no entanto, a entidade sem fins lucrativos pretende oferecer sua plataforma MOOC como um sistema de autosserviço, que os professores podem usar para desenvolver cursos especificamente de marca para suas universidades. A EdX, então, receberá os primeiros 50 mil dólares em receita gerada a partir do curso ou 10 mil dólares para um curso recorrente. Depois disso, a receita será dividida pela metade entre a universidade e a EdX. Um segundo modelo de geração de receita oferece às universidades “ajuda na produção” para o desenvolvimento de um curso, cobrando-lhes 250 mil dólares para um novo curso e 50 mil dólares para cada semestre em que ele for oferecido novamente. Se um curso for bem-sucedido, a universidade recebe 70% da receita, enquanto a EdX foi totalmente compensada por quaisquer cursos de autosserviço. No entanto, a fim de gerar receita suficiente para compartilhar com suas 12 universidades parceiras, que agora incluem Universidade da Califórnia, Berkeley, Wellesley, Georgetown e Universidade do Texas, um modelo provável é o de licenciamento. Testado sem custos em San Jose State University, em 2012, um MOOC da EdX serviu como base para um curso misto (on-line e presencial) de Engenharia. O currículo enriquecido resultou em um aumento da taxa de aprovação de 60% para 91%. Se o licenciamento do curso tornar-se o fluxo principal de receitas, Anant Agarwal, o engenheiro eletricitista e presidente da EdX, prevê que isso acontecerá em salas de aula lotadas com inscrições limitadas.

Mas alguns membros do corpo docente de San Jose State University estão, no entanto, alarmados. Quando um segundo MOOC da EdX, o JusticeX, foi considerado, o departamento de Filosofia enviou uma carta com palavras duras dirigidas ao desenvolvedor do curso de Harvard, Michael Sandel, mas que na verdade

tinham como alvo os administradores da universidade. Alegando que o departamento não tinha um problema acadêmico com necessidade de reforço e que não faltavam professores para ensinar o seu curso equivalente, não hesitou em atacar os motivos econômicos por trás da adoção de MOOCs pelas universidades públicas. Os autores alegaram ainda que os MOOCs representavam um declínio na qualidade educacional e observaram a ironia envolvida quando um curso de justiça social era o veículo para cometer uma injustiça social — um esforço de longo prazo para “desmantelar departamentos e substituir os professores”. A resposta conciliatória de Sandel expressou seu desejo de compartilhar recursos educacionais livres, sua aversão a subestimar colegas e uma convocação para um debate sério, tanto do pessoal da EdX quanto da comunidade de ensino superior.

Outras universidades estão semelhantemente posicionando-se contra, tanto à EdX quanto a outros novos empreendimentos MOOC tais como Coursera e Udacity, fundados por membros do corpo docente de Stanford. MOOCs e AES são indissociáveis. Os cursos on-line massivos necessitam de sistemas de avaliação automatizada. E Coursera e Udacity expressaram seu compromisso em usá-los, devido ao valor do feedback imediato. Os docentes da Faculdade Amherst votaram contra aderir ao consórcio EdX. Os membros do corpo docente da Universidade Duke frustraram as tentativas da administração para se juntar a outras nove universidades e à empresa 2U de tecnologia educacional em um empreendimento para desenvolver uma coleção de cursos de graduação para obtenção de crédito.

Mas a EdX foi fundada por duas das mais proeminentes universidades nos Estados Unidos, reuniu parceiros prestigiados e já está moldando os padrões educacionais. Stanford, por exemplo, decidiu apoiar; ela adotou a plataforma de código aberto OpenEdX e começou a oferecer um programa de preleção de verão para calouros e dois cursos públicos no verão de 2013. Stanford colaborará com a EdX no desenvolvimento futuro de OpenEdX, e oferecerá cursos públicos e universitários através dela.

Assim, enquanto o professor Perelman brinca que seus ex-alunos de Ciência da Computação poderiam desenvolver um app Android capaz de produzir dissertações estereotipadas que obteriam uma nota seis do *e-Rater*, eliminando completamente os seres humanos da equação, ele sabe que questões sérias estão em jogo. Quais resultados educacionais resultarão da diminuição da interação humana e da entrada efetuada por seres humanos? Será que a IA se desenvolverá até o ponto em que a verdade, o rigor, a organização eficaz, o poder de persuasão, a argumentação e os elementos comprovantes possam ser avaliados? E quantos outros postos de trabalho na educação ainda desaparecerão como resultado?

Fontes: Michael Gonchar, "How Would You Feel About a Computer Grading Your Essays?", *New York Times*, 5 abr. 2013; Steve Kolowich, "How EdX Plans to Earn, and Share, Revenue From Its Free Online Courses", *The Chronicle of Higher Education*, 21 fev. 2013; John Markoff, "Essay-Grading Software Offers Professors a Break", *New York Times*, 4 abr. 2013; Ry Rivard, "Humans Fight Over Robo-Readers", *Inside Higher Ed*, 15 mar. 2013; David Rotman, "How Technology Is Destroying Jobs", *MIT Technology Review*, 12 jun. 2013; Randall Stross,

"The Algorithm Didn't Like My Essay", *New York Times*, 9 jun. 2012; Michael Winerip, "Facing a Robo-Grader? Just Keep Obfuscating Mellifluously", *New York Times*, 22 abr. 2012; Paul Wiseman, Bernard Condon, e Jonathan Fahey, "Can smart machines take your job? Middle class jobs increasingly being replaced by technology", *The Associated Press*, 24 jan. 2013; "San Jose State University Faculty Pushes Back Against EdX", *Inside Higher Ed*, 3 mai. 2013.

PERGUNTAS SOBRE O ESTUDO DE CASO

- 11.13 Quão "inteligente" é a avaliação automatizada de dissertação? Justifique sua resposta.
- 11.14 Quão eficaz é essa avaliação?
- 11.15 Quais os benefícios da avaliação automatizada de dissertação? Quais as desvantagens?
- 11.16 Quais fatores humanos, organizacionais e tecnológicos devem ser considerados na decisão de utilizar a AES?
- 11.17 Você ficaria desconfiado de uma nota baixa obtida em um artigo avaliado pelo software AES? Justifique. Você solicitaria uma revisão por parte de um avaliador humano?