

O conceito *big data* na educação

Pasqueline Dantas Scaico^{1,2}, Ruy José G. B. de Queiroz¹, Alexandre Scaico²

¹ Centro de Informática (CIn) – Universidade Federal de Pernambuco

² Centro de Ciências Aplicadas e Educação (CCAIE) – Universidade Federal da Paraíba

{pds, ruy}@cin.ufpe.br, alexandre@dce.ufpb.br

Abstract. *Nowadays the Communication and Information Technologies have been widely used as a support platform to the learning process. A large amount of data related to this process can be automatically generated and stored to further analysis. As these data can be huge a technique that rises as an appropriate to analyze these data is called big data. In this paper we will present the big data concept emphasizing it uses in the learning process. We will present its benefits, actual uses and the challenges of its implementation in the educational context. What we intend is to start a discussion about how this data processing technique has been using as a support to the improvement of the learning process.*

Resumo. *Nos dias atuais as tecnologias de informação e comunicação (TIC) vêm sendo amplamente utilizadas como plataforma de apoio ao processo de ensino-aprendizagem. Uma grande quantidade de dados sobre esse processo pode ser gerada automaticamente e armazenada para análise. Como esses dados podem ser de grande volume, uma técnica que desponta como apropriada para análise é a chamada big data. Neste artigo apresentamos o conceito de big data, com ênfase em seu uso nos processo de ensino-aprendizagem. Serão apresentados seus benefícios, usos atuais e desafios em sua implementação no contexto da educação. Com isso pretende-se divulgar essa tendência para que seja possível iniciar futuras discussões sobre o uso da técnica como apoio à melhoria do processo de ensino-aprendizagem no Brasil.*

1. Introdução

Assim como ocorre com as atividades de comércio, entretenimento e comunicação, cada vez mais o processo educacional tem utilizado as Tecnologias de Informação e Comunicação (TIC) como uma plataforma de apoio ao processo de aprendizagem, e tem sido gerado uma grande quantidade de dados sobre esse processo.

O conceito de *big data*¹ está relacionado à capacidade de processar e analisar grandes volumes de informação que permitam a extração de conhecimentos úteis para melhorar o processo de tomada de decisão (US Department of Education, 2012). No campo educacional, as técnicas relacionadas ao *big data* podem tornar mais eficazes os processos de avaliação, feedback e entrega do conteúdo. Essa prática tem gerado a possibilidade de capturar muitas informações sobre os estudantes e sobre a sua interação com os conteúdos instrucionais, ambientes de aprendizagem e sobre o processo de avaliação, que são difíceis de ser coletadas e interpretadas por meio de abordagens tradicionais que não façam uso de tecnologias com alto poder de processamento e análise.

A utilização de *big data* foi mais uma vez apontado no Horizon Report de 2014 como uma área de pesquisa de alto impacto para educação do futuro, dado o seu grande potencial para o suporte à construção de ambientes adaptativos de aprendizagem que sejam capazes de responder às demandas dos estudantes em tempo real e promover maiores níveis de engajamento (Johnson et al, 2014). Técnicas específicas para manipulação de *big data*, já bastante utilizadas na área de negócios, têm começado a ser utilizadas no campo educacional na tentativa de entender o comportamento e os interesses dos estudantes e os fatores que podem levá-los a níveis maiores de engajamento.

O uso de *big data* na educação pode auxiliar a atuação de educadores, gestores e formuladores de políticas educacionais e, conseqüentemente, favorecer melhorias nos ganhos de aprendizagem. Todavia, é um conceito que precisa ser compreendido pelos estudiosos e interessados na área de Educação, especialmente para que novas soluções que utilizam a tecnologia possam ser projetadas tirando o máximo proveito do que podem fornecer os especialistas das mais distintas áreas: educação, psicologia e computação, por exemplo. Para tanto, é preciso conhecer primeiramente as tendências e as tecnologias que estão emergindo nesse cenário.

Este artigo caracteriza em linhas gerais os conceitos sobre *big data*, assim como, o seu potencial para a educação e as possibilidades que vêm sendo exploradas e que podem modificar no futuro o conhecimento científico existente sobre os processos de ensino e aprendizagem. Considerando que, atualmente, a maior parte da literatura existente sobre este tema tem sido produzida fora do país, este trabalho visa contribuir para a divulgação do conceito no país através de uma revisão bibliográfica, que ilustra o potencial do *big data* para melhorar o cenário educacional e os desafios que estão sendo enfrentados.

O artigo está organizado da seguinte maneira: a Seção 2 apresenta algumas reflexões sobre a importância de compreender o conceito de *big data* e especialmente o seu uso no campo educacional; na Seção 3 alguns exemplos são fornecidos para clarificar o potencial do uso de *big data* neste campo, assim como, os tipos de dados que têm sido objeto de interesse nas análises; a Seção 4 apresenta algumas aplicações efetivas do conceito de *big data*; na Seção 5 alguns desafios são pontuados, como forma de mostrar que ainda há muitas dificuldades da utilização plena de estratégias com *big data*; por fim, na Seção 6, são apresentadas as considerações finais.

¹ Os autores optaram pela utilização do termo em inglês.

2. *Big data* e educação

Analisar grandes volumes de dados pode apontar evidências, padrões e tendências que podem aprimorar e inovar estratégias de ensino e aprendizagem. Um exemplo de como o uso de grandes volumes de dados pode gerar impactos positivos na educação é o Programa Internacional de Avaliação de Alunos (*Programm for International Student Assessment - PISA*), que é organizado pela Organização para a Cooperação e Desenvolvimento Econômico e tem procurado analisar as habilidades de estudantes nas áreas de Matemática, Ciências e Leitura/interpretação, de forma que seja possível comparar as competências que os jovens estão desenvolvendo em diferentes países. O PISA tem disponibilizado instrumentos importantes para motivar os governos a entenderem como a alocação de recursos em outros países tem favorecido melhorias no campo educacional. As escolas também podem decidir entender o que as tornam tão semelhantes ou diferentes de outras em qualquer parte do mundo (Schleicher, 2013).

Dispor de mais conhecimentos sobre o processo de ensinar e aprender é importante para possibilitar que as instituições de ensino melhorem seus modelos de gestão de aprendizagem, de avaliação, construam novas práticas de ensino e discutam mais as políticas educacionais existentes. A utilização de *big data* na educação através do incentivo de governos, universidades e empresas pode estabelecer novas tecnologias, ferramentas e recursos que são capazes de apoiar uma cultura orientada ao conhecimento, à eficiência, à aprendizagem adaptativa e personalizada e promovedora de novas experiências de aprendizagem capazes de melhorar a maneira como os professores ensinam, os estudantes aprendem e a escola funciona.

O uso de *big data* pode reunir e analisar uma imensidade de dados que são produzidos pelas relações de ensino-aprendizagem. Com o uso de tecnologias pode-se coletar muito mais informações, em variados graus de granularidade, que os métodos manuais não são capazes de coletar. Capturar evidências que apontem a sequência de passos que um estudante usou para resolver um problema, suas estratégias, o número de conselhos que precisou receber e o tempo gasto na análise de cada parte do problema são alguns exemplos. A capacidade de processar massas de dados em escala, através da análise e da comparação de comportamento de milhares de estudantes, é muito importante para gerar conhecimentos generalizáveis sobre o processo de aprender e as dificuldades inerentes a determinados conteúdos ou estratégias pedagógicas.

Além disso, quando um estudante está envolvido em uma atividade de aprendizagem é importante que ele possa receber estímulos compatíveis com o seu comportamento, seu nível de proficiência e dificuldades. Isto significa que se o processo de avaliação funcionar em paralelo ao processo instrucional será possível produzir a tempo informações apropriadas para estabelecer os próximos estímulos. Um ciclo de feedback contínuo, que é capaz de operar em diferentes escalas de tempo e para diferentes atores, é importante para permitir que o estudante aprenda mais rapidamente; para que o professor visualize a evolução de aprendizagem dos seus estudantes e da turma em menores intervalos de tempo, permitindo-lhe planejar adequadamente o próximo encontro; que uma base de informações que agregue valor à avaliação dos estudantes e das práticas pedagógicas em uso seja estabelecida e que as instituições de ensino construam uma visão sistêmica sobre o desempenho geral dos seus alunos e das

práticas institucionais, algo que pode ser de grande utilidade para orientar mudanças e melhorias em maior escala em toda a escola.

3. Onde a educação se beneficia com o uso de *big data*?

O uso de *big data* pode ser a base para o desenvolvimento de algoritmos e aplicações que sejam capazes de criar inferências sobre o conhecimento que os estudantes possuem, suas intenções, interesses, assim como, a construção de modelos capazes de prever comportamentos e interesses futuros (Manyika et al, 2011). Na educação, este conceito pode auxiliar a responder questões complexas, como: o que um estudante sabe? O que ele não sabe? Até que ponto um conteúdo ou uma estratégia pedagógica pode funcionar para um grupo de estudantes? Que sequência de assuntos pode ser mais adequada para um estudante com um perfil específico? Que resultados de aprendizagem podem ser medidos quando um estudante interage com um determinado tipo de conteúdo? Até que ponto uma avaliação realmente atingiu a medição a que se propunha? Como ter certeza sobre o nível de proficiência de um aluno sobre um determinado conteúdo curricular? A falha do aluno ocorreu porque ele não aprendeu ou ela se deveu à distração, esquecimento ou a uma pergunta mal formulada? Como aumentar as chances de um estudante obter sucesso em uma avaliação?

Especialmente em se tratando de individualização das estratégias de ensino em função de necessidades específicas de cada aluno ou entregar estímulos adequados ou decidir que problemas devem ser resolvidos por cada estudante é importante dispor de conhecimentos que atestem o que um estudante é capaz de fazer e o que ele aprendeu. Estabelecer esse nível de inferência requer a observação de dados individuais acumulados e a comparação entre inúmeros desses dados. Por outro lado, ser capaz de classificar e agrupar estudantes que possuem o mesmo perfil e habilidades ou que enfrentam as mesmas dificuldades é uma maneira eficiente de articular algumas atividades e intervenções pedagógicas.

Diversos comportamentos dos estudantes podem ser tratados para melhorar o que se conhece sobre a relação de ensinar e aprender. Estudar o nível de persistência de um aluno que busca alternativas para resolver problemas complexos, a sua disposição para completar determinadas tarefas ou até mesmo o seu nível de assiduidade e de atrasos podem ser alguns exemplos. A identificação de comportamentos que ocorreram no passado e previsão de comportamentos que se espera no futuro pode também auxiliar a condução de estratégias pedagógicas mais eficientes que estimulem ações para a melhoria da aprendizagem, do engajamento e da retenção dos alunos.

A interpretação de grandes volumes de dados também pode conduzir mais eficientemente o planejamento de cursos, apontando como um conteúdo pode ser melhor dividido em tópicos, como estes devem ser sequenciados e que configurações podem gerar curvas maiores de aprendizagem. Outra aplicação do conceito é a análise de tendências, que pode apontar os impactos ocasionados pelas mudanças para o ambiente escolar como um todo. Dessa forma, o uso de técnicas de mineração e análise de volumosas bases de dados educacionais provenientes, por exemplo, de sistemas tutores inteligentes pode descobrir padrões, desenvolver algoritmos e modelos capazes

de apontar elementos sobre o comportamento e sobre aquilo que os estudantes aprenderam. Por outro lado, analisar os dados provenientes das ações dos próprios gestores educacionais e dos produtos resultantes do processo de ensino-aprendizagem podem gerar novos conhecimentos que influenciarão diretamente a prática educacional.

De acordo com (Ferreira, 2013), os tipos de informações que têm sido objeto de interesse na coleta e análise das aplicações que processam grandes volumes de dados educacionais dizem respeito aos:

- a) Dados relacionados à identidade dos usuários: quem são eles, quais são as suas permissões e de onde o acesso está sendo realizado;
- b) Dados relacionados à interação do usuário com o ambiente de aprendizagem: se referem à experiência que os estudantes tiveram ao utilizar o ambiente. Dados dessa categoria incluem métricas de retenção, taxa de cliques, páginas visitadas e taxas de retorno;
- c) Dados inferidos sobre os conteúdos: se referem à compreensão de evidências que sejam capazes de apontar o nível de eficiência existente na relação entre um conteúdo escolar e os ganhos de aprendizagem de estudantes ou grupos de estudantes. Alcançar métricas para gerar este tipo de dados é mais difícil por exigir a exposição de grupos de estudantes a diferentes conteúdos e posterior avaliação, que pode resultar em evidências que apontem a eficiência dos conteúdos;
- d) Dados relacionados aos sistemas de informação disponíveis nas instituições de ensino: se referem aos dados de natureza mais burocrática, que já são, em grande parte, coletados pelos mecanismos tradicionais da escola. Envolve notas, registros disciplinares e de assiduidade. Em pequena escala, este tipo de dados é de pouca utilidade, já que há pouca informação dessa natureza por cada estudante. Em larga escala os dados podem assistir à geração de recomendações para a escola como um todo, apesar de ainda se constituírem dados superficiais que podem não ser confiáveis ou de pouca utilidade para a tomada de ações, especialmente pelo fato de apontarem problemas sociais difíceis de serem resolvidos no contexto escolar;
- e) dados de inferência sobre os estudantes: são os dados mais difíceis de serem gerados porque dizem respeito ao comportamento de cada aluno diante do processo de aprendizagem de um conteúdo, que pode ser afetado em diferentes momentos por diferentes fatores e pelos recursos instrucionais utilizados. Dados dessa natureza procuram inferir o que um aluno aprendeu, qual é o seu comportamento e como isso influencia a sua aprendizagem, quais são as suas motivações e interesses, dentre outros aspectos.

4. Quem está aplicando big data na educação?

Arnold e Pistilli (2012) relatam que, de 2007 a 2010 a universidade de Purdue conduziu um projeto chamado Course Signal que fazia uso de *big data* para reunir conhecimento sobre seus estudantes a partir de diferentes fontes, com o intuito de gerar intervenções pedagógicas e identificar alunos em situação de risco, os quais precisavam

receber mais suporte. Ao que parece, o conhecimento que vem sendo descoberto tem sido importante para combater a evasão escolar, por exemplo, na instituição. A universidade comunitária de Saddleback também já utilizou um sistema de recomendação para personalizar a assistência aos seus alunos, chamado de SHERPA, que agrega dados dos estudantes (desde aqueles relacionados aos horários de trabalho até as suas experiências com professores) para construir recomendações que os orientassem sobre técnicas de gerência de tempo, seleção de cursos e outros fatores que pudessem contribuir com o sucesso acadêmico (Soares, 2012).

No final de 2012, a Course Smart², empresa fornecedora de livros digitais, anunciou um pacote de produtos, capaz de rastrear a interação dos alunos pelos textos digitais, algo que aumentava a capacidade de avaliação dos professores já que poderiam compreender com mais precisão o esforço individual dos estudantes e as suas decisões no momento de selecionar os textos. Outras plataformas comerciais de aprendizagem adaptativa têm utilizado o conceito de *big data*. É o caso da DreamBox³ e Knewton⁴.

Recentemente, o projeto inBloom⁵, com financiamento da Fundação Gates apresentou como missão o uso de *Big data* na implementação de um projeto de grandes proporções para transformar o ensino personalizado em algo possível. Algo que já vem sendo feito pela plataforma de ensino Khan Academy, que recolhe diariamente mais de oito milhões de pontos de dados. A plataforma é capaz de identificar por quanto tempo os vídeos foram assistidos, quantas tentativas foram necessárias para a resolução de um exercício e quanto tempo os alunos ficaram conectados à plataforma. Segundo Fauss (2014), o Khan Academy vem trabalhando em um modelo de previsão de comportamento, no qual será possível avaliar mais precisamente os ganhos de aprendizagem e que peças de conteúdo devem ser sugeridas individualmente para cada estudante. Essa escala de análise tem sido capaz de gerar mais significado para os dados, enquanto alunos e professores utilizam o Khan Academy.

Os cursos online e massivos (do inglês, Massive Open Online Courses – MOOC) também têm utilizado tais estratégias para saber mais sobre os estudantes. A plataforma Udacity, por exemplo, tem realizado alguns experimentos com a análise através de *big data* para identificar as variáveis que estão relacionadas ao comprometimento (ou não) dos estudantes em cursos introdutórios disponibilizados na plataforma (Collins, 2013).

5. Desafios

Apesar do processamento de grandes volumes de dados ser praticamente todo automatizado, a participação dos seres humanos é fundamental. Isto ocorre, por exemplo, no processo de análise visual dos dados, que ajuda a criar mais sentido para os mesmos ou na detecção inicial de padrões. A construção de sofisticados mecanismos de visualização que possam apresentar os dados de uma forma que as pessoas possam agir

² http://www.bizjournals.com/prnewswire/press_releases/2012/11/07/NY07704

³ <http://www.dreambox.com>

⁴ <http://www.knewton.com>

⁵ <https://www.inbloom.org>

de acordo com o aquilo que são capazes de interpretar também é uma atividade complexa.

Questões de padronização das informações e da interoperabilidade dos sistemas que processarão os dados ainda precisam ser discutidas para que se possa garantir a escalabilidade do conhecimento que está sendo produzido. Além disso, os custos de produção de ferramentas robustas que sejam capazes de lidar com o processamento dos dados nestas dimensões ainda são muito altos, assim como a formação de especialistas para trabalhar em uma área que exige a compreensão de conceitos relacionados à Computação, Psicologia, Estatística, dentre outras, conforme aponta Guthrie (2013). Inúmeras questões éticas precisam ser discutidas, para que seja possível preservar a privacidade dos dados de estudantes e professores, como menciona Kamenetz (2013). Este também é um outro assunto que precisa ser posto no centro das discussões.

6. Considerações finais

Este artigo apresentou um cenário geral sobre o conceito de *big data* na educação e o seu potencial disruptor para contribuir com mudanças criativas e significativas para este campo. A relevância deste conceito está no fato que ele representa um instrumento para que os tomadores de decisão, em diferentes níveis e esferas, possam ter os recursos necessários para analisar e prever o comportamento através de uma enorme diversidade de variáveis, que incluem: estilos de aprendizagem, dados geográficos e sociais, processos cognitivos, níveis de QI, intenções e objetivos de aprendizagem, fatores ambientes e até, pré-disposição genética, conforme menciona Briggs (2014).

Para o campo educacional esse nível de processamento de dados pode ser o suporte para envolver aspectos fundamentais para que possamos compreender melhor o que influencia e o que gera interferência nos processos de ensino e aprendizagem. Em primeiro lugar, a possibilidade de generalizar conhecimentos sobre um processo de aprendizagem aumenta se considerarmos que milhares de estudantes, por exemplo, recebendo as mesmas instruções, podem ter suas experiências individuais interpretadas de uma forma que atinja esse entendimento sistêmico que é necessário sobre o processo. Em segundo lugar, sabendo-se que a aprendizagem é mais efetiva quando o estudante pode dispor de um mecanismo de feedback instantâneo, algo que é quase impossível de ocorrer sem o uso da tecnologia, a velocidade no provimento de feedback e de novos estímulos para o estudante que está engajado em uma atividade é crucial para os seus resultados de aprendizagem, assim como, os rápidos ajustes que são necessários para melhorar a entrega do conteúdo e a própria experiência do estudante, que pode passar a ser individualizada.

Todavia, a extração de conhecimento que ocorre com a análise das bases de dados e com a tentativa de compreender determinados comportamentos, assim como a própria onipresença da tecnologia no ambiente educacional, têm levantado a uma discussão sobre inúmeras questões éticas. A extração e a manipulação de dados impõe a necessidade de regulação de ações para que se possa manter os princípios da privacidade, confidencialidade e transparência. Há ainda que aponte que o processo desumaniza a relação, reduzindo as análises apenas a uma perspectiva meramente

quantitativa do que parece representar o ensinar e aprender. Além do mais, a tendência em adotar cada vez mais avaliações de natureza somativa pode ser uma estratégia desmotivador para os estudantes e uma força negativa para os processos gerais de avaliação.

Para o cenário brasileiro, é importante que os pesquisadores na área, assim como os professores e tomadores de decisão, tomem conhecimento sobre as tecnologias que estão emergindo com esse novo poder de processamento, seus benefícios e prejuízos porque eles podem refletir uma nova cultura, práticas e comportamentos em relação aos processos de ensino e aprendizagem. Assim, este artigo se constitui um documento importante para que seja possível introduzir os conceitos associados a esse tema.

Agradecimentos

Este trabalho foi realizado no contexto do projeto “Um estudo sobre aprendizagem na era digital e o uso efetivo das tecnologias na escola”, que está sendo desenvolvido no Centro de Informática, Universidade Federal de Pernambuco. Os autores agradecem o suporte do CNPq (Proc. 486307/2013-1) a este projeto de pesquisa.

Referências

Arnold, K. E., Pistilli, M. D. (2012) Course Signals at Purdue: Using Learning Analytics to Increase Student Success. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pages 267-270, Vancouver, Canada.

Briggs, S. (2014) Big Data in Education: Big Potential or Big Mistake? Disponível em: <http://www.opencolleges.edu.au/informed/features/big-data-big-potential-or-big-mistake>. Último acesso: julho de 2014.

Collins, E.D., 2013. Preliminary summary: a study of Augmented Online Learning Environments, San Jose. Available at: http://www.sjsu.edu/chemistry/People/Faculty/Collins_Research_Page/AOLE_Report-September_10_2013_final.pdf. Último acesso: julho de 2014.

Faus, M. (2014). Big Data at Khan Academy. Blog. Disponível em: <http://mattfaus.com/2013/10/big-data-at-khan-academy>. Último acesso: julho de 2014.

Ferreira, J. (2013) Big data in education: The five types that matter. Disponível em <http://www.knewton.com/blog/knewton/from-jose/2013/07/18/big-data-in-education>

Guthrie, D. (2013). The Coming Big data Education Revolution. Disponível em: <http://www.usnews.com/opinion/articles/2013/08/15/why-big-data-not-moocs-will-revolutionize-education>. Último acesso: outubro de 2013.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A. (2014). NMC Horizon Report: 2014 Higher Education Edition. Austin, Texas: The New Media Consortium. Disponível em <http://cdn.nmc.org/media/2014-nmc-horizon-report-he-EN-SC.pdf> Último acesso: julho de 2014.

Kamenetz, A. (2013) How NSA Data Mining Shenanigans Could Sink InBloom And Big data In Education. Disponível em <http://www.wiredacademic.com/2013/07/how-nsa-data-mining-concerns-are-impacting-inbloom-and-big-data-in-education>. Último acesso: novembro de 2013.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. (2011). Big data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute. Disponível em http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation. Último acesso: agosto de 2013.

Soares, L. (2012) The Rise of Big Data. EDUCAUSE Review, vol. 47, no. 3. Disponível em <http://www.educause.edu/ero/article/rise-big-data>. Último acesso: julho de 2014.

Schleicher, A. (2013) Big data and PISA. Disponível em http://www.huffingtonpost.com/andreas-schleicher/big-data-and-pisa_b_3633558.html. Último acesso: novembro de 2013.

US Department of Education (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, Washington, D.C., Disponível em <http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf>. Último acesso: novembro de 2013.