Desafios para os Avanços da Análise de *Big Data* na Saúde

Adriana Benício Galvão¹, Ricardo Alexandre de Medeiros Valentim¹

¹Laboratório de Inovação Tecnológica em Saúde – Universidade Federal do Rio Grande do Norte (UFRN) – Natal, RN– Brasil

adriana.benicio@lais.huol.ufrn.br, ricardo.valentim@ufrnet.br

Abstract. *This paper describes challenges to the advances of Big Data Analysis in Health, which are relevant because of the current needs for high demand health systems. Therefore, it is important to develop new analytical tools that are easy to use, reflect the aggregation of knowledge from different areas and solve real health problems. As a way of validating the new tools, it is suggested to use data from the various Health Information Systems of the Brazilian Ministry of Health.*

Resumo. *Este artigo descreve desafios para os avanços da Análise de Big Data na Saúde, os quais se mostram pertinentes devido às necessidades atuais por sistemas de saúde de alta demanda. Diante disso, é apresentada a importância do desenvolvimento de novas ferramentas de análise, que sejam de fácil uso, reflitam a agregação de conhecimentos de diversas áreas e resolvam problemas reais de saúde. Como forma de validar as novas ferramentas, é sugerida a utilização de dados dos vários Sistemas de Informação em Saúde do Ministério da Saúde do Brasil.*

1. Introdução

Nos dias atuais, um dos maiores custos no mundo está nos investimentos em saúde pública. Fatores como o envelhecimento da população, o aparecimento precoce de doenças, a carência de cuidados preventivos, entre outros, aumentam as demandas por serviços de saúde e exigem deles maior eficiência. Para tanto, é importante destacar que os sistemas e serviços de saúde em todo mundo passam por um conjunto significativo de mudanças guiadas por uma nova tendência global, que vem sendo tratada como a “quarta revolução industrial”. Por trás desse conceito, está essencialmente concretada a sinergia global da convergência tecnológica, a qual tem relação muito intrínseca com os processos de automatização da indústria. A maior expressão desse novo paradigma parece estar se viabilizando por meio de sucessivos experimentos, estudos e pesquisas que criam laços de intercessões em diversas fronteiras do conhecimento [Ministério da Saúde 2018; Thuemmler 2017; Valentim et al. 2012]

Diante da tendência em interconectar conhecimentos de diversas áreas, são necessários avanços tecnológicos que possibilitem a comunicação e integração de dados entre diversos Sistemas de Informação em Saúde (SIS), além do processamento, exploração e análise do grande volume de dados produzidos, no intuito de extrair conhecimento e *insights* para formular ações que gerem resultados. Tais características

estão associadas ao conceito de Ciência dos Dados, o qual envolve a busca por padrões sem que as questões de negócio estejam necessariamente bem formuladas, diferentemente da análise tradicional, que identifica os dados que atendem determinados padrões a fim de resolver questões bem definidas. Essa diferença reflete um dos objetivos da análise de dados que é prever o que os dados podem representar no futuro, em vez de somente explicar o que significam. Nesse sentido, a Ciência dos Dados tem um potencial de mostrar meios de resolver e evitar problemas de saúde em geral, além de revelar informações pertinentes para o aprimoramento dos serviços prestados à população.

2. Descrição do Problema

No contexto de *big data*, em que há grande volume de dados de diferentes origens, em diferentes formatos, e que precisam ser processados em alta velocidade, algumas técnicas e ferramentas tradicionais apresentam desempenho insuficiente, demonstrando a necessidade de utilização de tecnologias específicas para o contexto. O ecossistema de tecnologias para *big data* apresenta uma variedade de ferramentas disponíveis. No entanto, o uso efetivo delas é pouquíssimo significativo na área da saúde. Há a necessidade atual de ferramentas facilitadoras para a análise de *big data* na saúde, em especial na Epidemiologia, que possam contribuir para a descoberta de novas formas de ação preventiva de doenças, padrões de comportamento epidemiológicos, correlações, periodicidades, intervenções não programadas e seus efeitos, para auxiliar nas tomadas de decisões, por meio da comparação de dados de fontes diversas, da conexão entre sistemas de informação de saúde e de uma infra-estrutura que permite escalabilidade.

A complexidade dos sistemas analíticos de *big data* é um dos fatores que implica no uso limitado das tecnologias envolvidas. Por isso, é importante desenvolver ferramentas de análise de *big data* que sejam de fácil utilização por estudiosos das questões de saúde pública. Para Jokonya (2014), o problema está na ausência de analistas habilitados em *big data* com conhecimento de saúde, que tenham a capacidade de identificar dados e ferramentas corretos para a análise de dados e interpretar as percepções obtidas. Outras possíveis causas-raiz dos problemas citados, que estão dentro do escopo da proposta deste artigo, incluem: a falta de agregação entre o conhecimento dos estudiosos nas áreas da saúde e da tecnologia, de integração entre bases de dados de Sistemas de Informação em Saúde e de disponibilização de técnicas tradicionais em ferramentas de análise de *big data*.

3. Relevância

A crescente quantidade de dados registrados eletronicamente todos os dias tem oferecido oportunidades para a geração de valor de negócio, melhor prestação de serviços à população e redução de custos em diversas áreas, especialmente na saúde pública, por meio da análise cuidadosa dos dados. James et al. (2011) acredita que *big data* pode ajudar a reduzir o desperdício e ineficiência na saúde pública por meio da análise de padrões de doença, rastreamento de surtos e transmissão para melhorar a vigilância da saúde pública e acelerar a resposta; do desenvolvimento mais rápido de

vacinas mais direcionadas; e da transformação de grandes quantidades de dados em informações acionáveis que possam ser usadas para identificar necessidades, fornecer serviços e prever e prevenir crises, especialmente para o benefício da população. Coelho Neto (2019) acredita que soluções de *big data* podem contribuir para a integração e comunicação dos vários sistemas, de sistemas de pequeno, médio e grande porte, maior capacidade de reuso de sistemas e diminuição de custos.

Não obstante, conforme Mehta & Pandit (2018), não há evidências mínimas sobre como a Análise de *Big Data* pode melhorar a qualidade dos cuidados na saúde e não há estudos de avaliação econômica sobre sua relação custo-benefício. Nesse sentido, é preciso incentivar o surgimento de estudos que mostrem a viabilidade de *big data* na saúde, por meio da realização de avaliações quantitativas, de trabalhos que utilizam dados reais e gerem benefícios reais, e do desenvolvimento de ferramentas simples e integradas que permitam que analistas possam realizar suas análises com foco nas questões de negócio.

4. Público alvo

O público que utilizaria possíveis ferramentas de análise de grande volume de dados, desenvolvidas na tentativa de solucionar os problemas expostos, seria composto por profissionais da saúde em geral, como epidemiologistas, médicos, cientistas de dados da saúde, etc. Os principais benefícios, entretanto, são direcionados à sociedade, à população em geral, pois são as necessidades sociais que estão constantemente demandando mudanças nas relações entre a tecnologia, profissionais de saúde e pacientes. Conforme Morais (2015), as mudanças na saúde global para o novo paradigma social centrado na quarta revolução industrial estão ocorrendo em partes pelo desenvolvimento e avanços tecnológicos, e em partes por pressão da própria sociedade, que todos os dias demanda por mais serviços de saúde, por melhores medicamentos, vacinas, por um melhor cuidado individual do paciente.

5. Soluções Existentes

A tabela 1 mostra estudos que buscaram conectar dados epidemiológicos e sistemas e gerar informações pertinentes para a geração de conhecimento em saúde.

Tabela 1. Artigos que buscaram conectar dados epidemiológicos e gerar conhecimentos em saúde

Artigo	Estudo desenvolvido
Google-driven search for <i>big data</i> in autoimmune geoepidemiology: Analysis of 394,827 patients with systemic autoimmune diseases, Ramos-Casals (2015)	Utilizou o motor de busca da Google para coletar e mesclar grandes séries de doenças autoimunes sistêmicas, com o objetivo de obter a figura geoepidemiológica de alta de definição de cada doença.

Leveraging hospital <i>big data</i> to monitor flu epidemics, Bouzillé & Chazard (2018)	Mostrou que grandes dados de hospitais parecem ter um grande potencial para monitorar epidemias de influenza, além de fornecer características adicionais dos pacientes.
Cloud enabled data analytics and visualization framework for health-shocks prediction, Mahmud et al. (2016)	Apresenta um framework de visualização e análise de dados para predição de surtos baseado em um dataset de larga escala com dados de 1000 lares no Paquistão.
Real-time predictive seasonal influenza model in Catalonia, Spain Basile (2018)	Projeta um modelo preditivo espacial em tempo real da taxa de incidência de ILI (Influenza Like Illness) na Catalunha usando previsões de uma e duas semanas, levando em consideração a distribuição geográfica dos surtos durante a temporada de influenza de 2015-2016.

Como iniciativas do governo brasileiro, no sentido de prevenir doenças, reduzir custos e ao mesmo tempo incentivar a formação de conhecimento tecnológico atualizado, pode-se citar o projeto Sífilis Não, o qual induz ações voltadas ao controle da sífilis nas redes de atenção à saúde, com atuação de apoiadores locais, investindo em pesquisas e na produção de conhecimentos por meio de estudos operacionais, ao passo que potencializa a capacidade técnica de vigilância e assistência local. Outro exemplo, que revela um pouco sobre o potencial do cruzamento de dados de diversas bases, está no aumento dos diagnósticos de HIV no Brasil, que teve um incremento de 11,3% em 1997, após decisão do MS em utilizar como critério de diagnóstico não só a notificação do Sinan net, mas também os dados do Siscel de CD4/CD8 e Carga Viral [Paz, Luiza e Dhalia, 2010]. Embora esse exemplo não seja relativo à análise de *big data*, ele consegue demonstrar que uma simples integração entre sistemas pode produzir um impacto positivo na produção de informação para a tomada de decisão.

6. Recursos para Propostas de Solução

Tanto para o desenvolvimento quanto para a avaliação das propostas de solução ou minimização dos desafios apresentados, seria importante a participação de profissionais de saúde que trabalham no planejamento de ações do governo para controle e prevenção de agravos de doenças, por exemplo. Quanto aos dados para análise, podem ser utilizados dados dos Sistemas de Informação em Saúde de base nacional, listados na tabela 2. A maioria deles está disponível no Tabnet: um tabulador de domínio público, desenvolvido pelo Datasus, que possibilita consultar as bases de dados do Sistema Único de Saúde (SUS). É importante ressaltar que, além dos dados listados na tabela 2, podem ser utilizados outros, tais como: informações demográficas e socioeconômicas,

dados do Cadastro Nacional de Estabelecimentos de Saúde (CNES) e do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP).

Tabela 2. Sistemas de Informação em Saúde e respectivos endereços para download do material de estudo

Sistema	Dados disponíveis e URL para acesso
CMD (Conjunto Mínimo de Dados)	Procedimentos realizados nas várias modalidades assistenciais: http://www2.datasus.gov.br/DATASUS/index.php?area=0202&id=33640385
Website de divulgações de informações do DIAHV (Departamento de Vigilância, Prevenção e Controle das IST, do HIV/Aids e das Hepatites Virais)	Boletins epidemiológicos referentes a AIDS/HIV, Sífilis e Hepatites Virais: http://www.aids.gov.br/pt-br/centrais-de-conteudos/boletins-epidemiologicos Painel de Indicadores Epidemiológicos: http://www.aids.gov.br/pt-br/gestores/painel-de-indicadores-epidemiologicos
SIH (Sistema de Informações Hospitalares)	Morbidade e Produção Hospitalar do SUS: http://www2.datasus.gov.br/DATASUS/index.php?area=0202&id=11633
SIA (Sistema de informações Ambulatoriais)	Produção ambulatorial: http://www2.datasus.gov.br/DATASUS/index.php?area=0202&id=19122
SIM (Sistema de Informações sobre Mortalidade)	Registro de óbitos: http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937
SINAN (Sistema de Informação de Agravos e Notificação)	Notificação de agravos compulsórios da Vigilância à Saúde: http://www2.datasus.gov.br/DATASUS/index.php?area=0203
SI-PNI (Sistema de informação do Programa Nacional de Imunização)	Imunizações: http://www2.datasus.gov.br/DATASUS/index.php?area=0202&id=11637

7. Resultados Esperados

O principal desafio exposto neste artigo está na criação de ferramentas que sejam utilizadas para realizar análises sobre dados de saúde pública, provenientes dos Sistemas de Informação do SUS. Como propostas de solução, são esperados protótipos de ferramentas que executem uma ou mais análises a respeito de doença(s) epidemiológica(s) recorrente(s), como AIDS/HIV, Sífilis, Hepatites Virais e Influenza A (H1N1). As soluções podem, por exemplo, estar associadas ao combate, prevenção, diagnóstico precoce, descoberta de padrões de comportamento, rastreamento e/ou previsão de doenças infecciosas. Dentre os critérios de avaliação dos protótipos, podem estar: potencial de agregação de valor das análises, facilidade de utilização da ferramenta por profissionais que não são da área de tecnologia, integração entre sistemas, quantidade estimada de dados envolvidos nas análises e inovação das soluções.

Referências

- Coelho Neto, Giliane Cardoso (2019). Integração entre Sistemas de Informação em Saúde: o caso do e-SUS Atenção Básica. Dissertação (Mestrado) – Universidade Federal de São Paulo. Escola Paulista de Medicina. Programa de Pós-Graduação em Saúde Coletiva. – São Paulo. xi, 104f.
- James, Manyika, Chui Michael, Brown Brad, Bughin Jacques, Richard Dobbs, Charles Roxburgh & Angela Hung Byers (2011), 'Big data: The next frontier for innovation, competition, and productivity', McKinsey Global Institute (June), 156.
- Jokonya, Osdan (2014), 'Towards a Big Data Framework for the Prevention and Control of HIV/AIDS, TB and Silicosis in the Mining Industry', *Procedia Technology* 16, 1533–1541.
- Mehta, Nishita & Anil Pandit (2018), 'Concurrence of big data analytics and healthcare: A systematic review', *International Journal of Medical Informatics* 114, 57–65.
- Ministério da Saúde (2018). Avanços, desafios e oportunidades no complexo industrial da saúde em serviços tecnológicos – Brasília: Ministério da Saúde. 308 p.
- Morais, Antonio Higor Freire (2015). Fuzzy angel: uma arquitetura distribuída de telemedicina para monitoramento de pacientes com esclerose lateral amiotrófica. 2015. 78 f. Tese (Doutorado em Engenharia Elétrica e de Computação) – Universidade Federal do Rio Grande do Norte, Natal.
- Paz, L. C.; Luiza, V. L.; Dhália, C. B. C. (2010). Avaliação da qualidade dos dados sistema de controle de exames laboratoriais (SISCEL) como fonte de identificação de casos de aids em crianças. *Cadernos de Saúde Coletiva*, Rio de Janeiro, v.18, p. 33-43.
- Thuemmler, Christoph; Bai, Chunxue (2017) *Health 4.0: How virtualization and big data are revolutionizing healthcare*. New York, NY: Springer.
- Valentim, Ricardo Alexsandro de Medeiros, et al. (2012) *Automação hospitalar: o estado da arte*. *Revista Brasileira de Inovação Tecnológica em Saúde*, v. 2.1.