

EDITORIAL

COVID-19: Desafios para modelagem e análise de dados

Edson Amaro Junior

*Médico Neurrorradiologista e Superintendente de Ciência de Dados e Analytics Hospital Israelita
Albert Einstein, São Paulo (SP), Brasil*

A utilização de tecnologias da informação para endereçar problemas de saúde é parte comum e já amplamente adotada. Entretanto a incorporação de ferramentas de predição, prescrição e otimização que evoluem rapidamente em complexidade ainda é bastante recente. Análises que permitem fazer predições relevantes a respeito de desfechos clínicos e em gestão de saúde são incorporadas na rotina operacional de uma pequena parte das instituições. A forma de trabalho das equipes que fazem uso dessas ferramentas também requer adaptações e deve focar no uso efetivo de seus resultados. E nesse contexto de desenvolvimento, vivemos uma pandemia histórica. As perguntas são mais emergentes, as respostas necessárias devem ser assertivas e este momento constitui um teste para os avanços em ferramentas analíticas: afinal, temos capacidade de usar estas tecnologias para enfrentar o surto de COVID-19? Como podem ser organizadas as iniciativas de análise de dados a fim de permitir predição e prescrição de cenários num futuro próximo baseado em passado recente?

A organização de trabalho em contextos de saúde pode permitir maior profundidade analítica, agregação de valores e respostas com um foco mais específico. Aqui é apresentada uma das formas de utilizar contexto de aplicação para organizar de equipes envolvidas em desenvolver: a) soluções preditivas de momento (D+0), para quando o paciente está em contato com agente de saúde (médico, enfermeira a equipe multi-profissional); b) soluções preditivas de curto prazo (D+15) gestão de saúde em população atendida por um hospital ou microrregião e c) soluções de ciclo epidemiológico, para o contexto de macro regiões envolvendo gestão populacional de um estado ou país. Ainda, nestas últimas abordagens incluir georreferenciamento nas predições (tempo-espço) adiciona grande valor para ações preventivas e de acompanhamento de casos.

Soluções em contexto de atendimento do paciente em geral envolvem predição de desfechos típicos no ciclo de pronto atendimento, enfermarias e UTIs. Questões podem ser respondidas com algoritmos de predição envolvem se o paciente está contaminado pelo SARS-COV2, se irá necessitar de um leito de UTI, se necessitará de assistência ventilatória ou diálise e qual será o desfecho mais provável ao final seu tratamento. Nessas soluções as ferramentas computacionais mais comuns envolvem redes melhores, árvores de decisão e eventualmente redes neurais profundas (quando se utilizam, por exemplo, informações de imagem e texto livre). A nossa experiência nesse cenário demonstrou ser possível desenvolver algoritmos com desempenho passível de adoção na rotina clínica, e que podem ser adotados em outros hospitais, considerando procedimentos de boas cartas de validação externa como primordiais nesse processo⁽¹⁻²⁾.

Soluções preditivas de curto prazo envolvem principalmente o planejamento da força de trabalho de profissionais em Hospitais ou microrregiões e também envolvem alocação de recursos “móveis” como equipamentos e re-utilização de espaços. As ferramentas analíticas utilizadas vão desde modelos de aprendizagem de máquina (supervisionados ou não) até técnicas estatísticas clássicas como projeção de demanda e análise de cenários. Também é possível a aplicação de ferramentas de otimização determinística (p.ex. simplex) ou probabilística (p.ex. computação evolutiva). As perguntas a serem respondidas envolvem planejamento de processos, por exemplo, para fluxo de atendimento de populações contaminadas e não contaminadas.

Soluções que envolvem análise da dinâmica epidemiológica em macro regiões são as mais desafiadoras. Envolvem perguntas tipicamente relacionadas à duração do ciclo epidemiológico, o número de pessoas infectadas, pessoas sintomáticas e a simulação de intervenções não farmacológicas que possam se tornar medidas determinadas pelos governos. Os dados utilizados envolvem grande variedade de fontes: movimentação espacial da população (p.ex. dados de telefonia móvel, GPS) questionários comportamentais (para avaliar aglomerações e atitudes de risco), dados de Clima/tempo, histórico de utilização de serviços de saúde, distribuição da população mais vulnerável, relação de fatores de risco sócio econômicos⁽³⁾ e de

deslocamento motivado (laboral, educacional ou recreativo). As ferramentas utilizadas são tipicamente os modelos clássicos de epidemiologia, dos quais modelos baseados em compartimentos (SIR, SIER e suas variações). Esses modelos são estudados há décadas e permitem adaptações, sendo importante lembrar de suas premissas de uso e da dependência de equações diferenciais. A integração de vários elementos analíticos em conjunto com desafios pela variedade de dados para esta abordagem, frequentemente trazem complexidade decorrente da dependência de inúmeros parâmetros cujas premissas de análise podem ser de difícil verificação. Ainda, estas ferramentas de predição de longo prazo em macro regiões são passíveis de grande variação inclusive pela sua dependência de fatores sem lastro em dados históricos, notadamente comportamento populacional e políticas públicas. Ainda, acrescenta-se nesta escala de complexidade as mutações virais⁽⁴⁾, frequência de diferentes cepas e a interação com outros patógenos.

É importante considerar que os contextos acima representam uma visão utilitária para tecnologias analíticas durante a pandemia. Certamente há várias outras propostas e com diferentes abordagens que podem contribuir na busca de melhor entendimento e apoio à decisão (análise de padrões de deslocamento entre cidades, referenciamento de pacientes no sistema de saúde por redes complexas e grafos, são alguns exemplos). Ainda, a qualificação de dados, a variabilidade populacional, desigualdade social e diferentes ações gestão pública tornaram desafiadora a utilização de apenas um modelo analítico para todos cenários. Acredito que este é um momento de frustração daqueles que esperavam respostas mais prontas e imediatas de algoritmos de aprendizagem de máquina recentes (notadamente redes neurais de aprendizagem profunda - *deep learning*), valendo lembrar da sua dependência de grande volume de dados – tipicamente ausentes no início de uma epidemia. Mesmo com a utilização de técnicas de transferência de aprendizado, por exemplo, os resultados de muitas destas abordagens não foram suficientes para que fossem adotados na rotina.

Por fim apesar de se tratar de momento difícil na história da humanidade, a necessidade de devolver soluções na velocidade necessária fez com que esse surto de COVID-19 proporcionasse grande avanço no desenvolvimento e utilização de técnicas analíticas avançadas. Elas contribuíram para abordar problemas de alta complexidade, com variados níveis de sucesso. É certamente esse esforço já trouxe benefícios e deverá acelerar a adoção de soluções digitais na saúde, assim esperamos.

REFERÊNCIAS

1. Deliberato RO, Escudero GG, Bulgarelli L, Neto AS, Ko SQ, Campos NS, et al. SEVERITAS_ An externally validated mortality prediction for critically ill patients in low and middle-income countries. *International journal of medical informatics*. Elsevier; 2019 Nov 1;131:103959.
2. Cosgriff CV, Celi LA, Ko S, Sundaresan T, la Hoz de MÁA, Kaufman AR, et al. Developing well-calibrated illness severity scores for decision support in the critically ill. *npj Digital Medicine*. Springer US; 2019 Aug 8;1-8.
3. Barrozo LV, Fornaciali M, de André CDS, Morais GAZ, Mansur G, Cabral-Miranda W, et al. GeoSES: A socioeconomic index for health and social research in Brazil. Lanza Queiroz B, editor. *PLoS ONE*. Public Library of Science; 2020;15(4):e0232074.
4. Candido DDS, Claro IM, de Jesus JG, Souza WM de, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *medRxiv*. Cold Spring Harbor Laboratory Press; 2020 Jun 11;:1-13.