

Instituto Federal de Pernambuco - UFPE

Centro de Informática - CIn

Princípios éticos e morais das ações tomadas por algoritmos

Douglas Soares Lins - dsl

Ramon de Saboya Gomes - rsg3

Recife, 2018.

Princípios éticos e morais das ações tomadas por algoritmos

A ciência dos dados oferece grandes oportunidades para melhorar e facilitar a vida privada e pública, como na saúde, utilizado para o cuidado de pacientes, avaliando melhor as suas condições e fornecendo recomendações que fortalecem a saúde e evitam despesas médicas futuras. Infelizmente, essas oportunidades estão associadas a desafios éticos, que diz respeito, por exemplo, a privacidade, discriminação e autonomia dos algoritmos.

Diversas organizações, como IEEE e ACM, se preocupam com os problemas relacionados a ética dos dados e possuem seus próprios códigos de conduta relacionadas a Big Data, apesar dos códigos não serem universais, são bastantes parecidos. Ética dos dados é o ramo da ética que estuda e avalia os problemas morais relacionados aos dados (geração, processamento, compartilhamento e uso), algoritmos (incluindo inteligência artificial, aprendizado de máquina e bots) e práticas correspondentes (incluindo inovação, programação e códigos profissionais), a fim de formular e apoiar soluções moralmente boas.

É necessário estruturar a discussão, visando identificar quais as preocupações envolvidas com os princípios éticos e morais da forma como os dados são utilizados nos algoritmos e das ações tomadas com base nesses dados. Também é importante ter a possibilidade de identificar problemas e as causas dos problemas envolvendo esses princípios (rastreadabilidade).

Existem as preocupações epistêmicas, que tratam do quão confiável são as conclusões obtidas pelos algoritmos. As evidências que levam a conclusão podem ser: inconclusivas (apesar de, no geral, ser estatisticamente suportado, o conhecimento pode ser incerto), inescrutável (impossível de ser entendido, isto é, não há como determinar como o algoritmo chegou ao conhecimento) ou equivocadas (por serem totalmente baseadas nos dados fornecidos, a qualidade da evidência está totalmente atrelada a qualidade dos dados).

As preocupações normativas tratam das ações tomadas baseado nas evidências produzidas pelos algoritmos e são: ações injustas e *transformative effects*. Mesmo com uma base de dados conclusiva e bem estruturada, uma ação direcionada pelas evidências de um algoritmo pode ser considerada discriminatória em alguma instância, gerando um resultado injusto. Ainda assim, as ações tomadas a partir das inferências de um algoritmo podem ser consideradas como justas ou corretas, pois podem não aparentar causar nenhum mal imediato.

Como as ações tomadas a partir de evidências obtidas com os algoritmos são sensíveis e podem afetar a vida de uma pessoa a curto ou longo prazo, há quem defenda que os *thresholds* e configurações gerais dos algoritmos devem ser definidas pelos usuários que irão utilizar o sistema,

já que eles detêm mais conhecimentos que os engenheiros de software que o produziram.

Os artigos focam em definir conceitos para auxiliar na discussão, assim como apresentam exemplos de abordagens e casos em que seja possível observar as diferenças entre essas abordagens. Também apresentam propostas de aprofundamento na literatura, devido ao tamanho do escopo do estudo desses algoritmos.

Apesar de apresentar grande potencial, utilizar um algoritmo para tomar uma decisão torna a situação complexa, pois existem casos raros e difíceis ou até mesmo impossíveis de programar. Estas situações podem ocorrer em qualquer tipo de situação, como na área de saúde, que tem em jogo a vida de uma pessoa. Este desafio tecnológico, é de extrema importância para garantir que desde o processamento das informações até as decisões feitas pelo algoritmo esteja de acordo com os valores éticos e morais impostos pela sociedade.