

Universidade Federal de Pernambuco

Centro de Informática

Curso de Ciência da Computação

Big data e suas consequências

Equipe:

Arthur Freitas (abf2)

Leonardo Espíndola (lre)

Outubro de 2018



O que é Big Data

Existem diversas definições do que seja Big Data, comumente é conhecido por descrever um amplo conjunto de dados que podem ser tanto estruturados quanto não estruturados que podem ser gerados em curtos períodos de tempos a todo instante. O termo pode ser novo, mas o conceito já existe há bastante tempo [1] porém, atualmente, com o advento dos gadgets, houve uma explosão exponencial da quantidade de informação, a velocidade como ela é produzida, a variedade de como ela é obtida, esses são os Big V's do Big Data. Ainda existem dois outros V's, Veracidade que está diretamente ligado ao quanto a informação coletada é verdadeira e Valor que seria o quão útil a informação obtida realmente é.

No que tange o conceito de Big Data, além da categorização de dados em relação a sua estrutura (estruturados e não-estruturados), ainda existem os conceitos para classificação dos dados que são obtidos a seu tipo, sendo eles Social Data, são dados oriundos das pessoas e tipos de informações que decifram comportamentos, como dados produzidos por comentários em redes sociais ou históricos de navegação. Há também os tipos de dados classificados como Enterprise Data, nos quais são os dados gerados por empresas a todo momento (dados do financeiro, recursos humanos, operações etc), estes podem ser usados para criação de indicadores de produtividade, fatores para tomada de decisão e possíveis gargalos. Os dados originados pela internet das coisas formou uma nova classificação chamada de Data of Things, esses são gerados a partir de informações que geladeiras, carros, TVs e outros dispositivos conectados a internet que conversam entre si.

Problemas de segurança relacionados a Big data e a ameaça à privacidade dos consumidores

O crescimento na geração de dados em relação a tudo, pode ser útil a sociedade para prever coisas importantes como o tempo, mas sofre riscos de privacidade e segurança. A quantidade de informação vazadas em um *breach* é gigante quando se trata de big data e é extremamente difícil de ser evitado pela forma como a mesma é armazenada, de forma distribuída. Uma das formas de manter informação segura é usar apenas um ponto de acesso, que é impossível com big data.

A consequência de se investir em uma maior segurança é que pode-se causar mais argumentos para coletar cada vez mais dados sensíveis. Por exemplo, as agências do governo podem requisitar às grandes companhias de big data informações sobre pessoas específicas, que

essas podem ser usadas em questões de segurança observando terroristas mas ao mesmo tempo para um motivo torpe como de vigiar pessoas específicas [2].

Big data pode ser utilizado para rastrear vários hábitos da população, percursos diários, sites visitados, até mesmo traçar personalidades. As empresas detentoras desse tipo de informação afirma que com essa informação é possível melhorar a experiência de usuários com recomendações adaptadas para cada um. Porém, também pode facilmente ser usada de forma a classificar usuários, na análise de crédito por exemplo.[3]

Transparência é chave para que se possa aproveitar a era de big data, é preciso que as empresas mostrem como os dados são coletados, para quais propósitos e como a informação é armazenada.

Análise Preditiva, privacidade e o caso da empresa Target

A Linha de supermercados Target necessitou conhecer seus clientes, decorrente disso, foi feito um trabalho de análise dos consumidores na qual associa-se um id para representar um usuário vinculado ao cartão de crédito usado, nome ou email e assim recolhe-se dados do usuário. Usando esses dados e os dados das mulheres que se cadastraram no pacote da Target para grávidas, notou-se um padrão em suas compras: com um crescimento na compra de itens como loção sem cheiro, sabão sem cheiro, desinfetante para mãos, etc. Através de 25 produtos o Target pode prever até mesmo quando o bebê iria nascer e pode utilizar isso como propaganda com foco nas pessoas que provavelmente terão um novo filho. Para encobrir o quanto a empresa sabia, passou-se a mandar propaganda sortida de itens para os novos pais com propaganda normal.[5][6]

Regulamentação de big data é muito difícil, para tentar controlá-la foi proposto o GDPR que foi implementado em maio de 2018. Ele visa estender o controle sobre empresas estrangeiras que tratam de dados de cidadãos estrangeiros e propõe uma regulação bem estrita que se não cumprida pode custar uma multa de 2% de toda faturação da empresa.[7]

Com big data é possível inferir dados como o exemplo da target, que fica bem difícil de regular, isso é chamado de análise preditiva. Independente da regulação sobre os dados que uma empresa recolhe, o uso dessa técnica pode acabar resultando na obtenção de informações privadas que tecnicamente não são recolhidas, logo não regulamentadas.

Outro possível dano do uso de análise preditiva, é que a informação obtida pode ser danosa ao usuário. Por tratar-se apenas de um chute educado a informação pode estar errada e o indivíduo ser julgado por isso.

Estado de vigilância na china

Não são só as empresas privadas que estão se utilizando de big data, governos como a china tem uma rede enorme de vigilância sobre o governo. De um nível que remete a instaurada na oceania no clássico 1984. As câmeras acompanham uma pessoa, pegam dados de seu rosto e a comparam a um grande banco de dados de forma a saber quem é a pessoa sendo vigiada. As câmeras auxiliam na procura de criminosos, porém representam um pesadelo aos cidadãos em busca de privacidade como na obra de george orwell. O sistema pode guardar até uma semana de comportamento de uma pessoa, e de fato já está sendo utilizado para expor pedestres que atravessam fora da faixa e limitar até mesmo o uso de papel higiênico.[8]

Referências

- [1]<https://www.winshuttle.com/big-data-timeline/>
- [2]<https://edition.cnn.com/2013/09/27/politics/nsa-snooping/>
- [3]<https://theconversation.com/big-data-security-problems-threaten-consumers-privacy-54798>
- [4]<https://dataconomy.com/2014/11/how-big-data-is-changing-the-insurance-industry-forever/>
- [5]http://www.slate.com/articles/technology/future_tense/2014/09/data_use_regulation_the_libertarian_push_behind_a_new_take_on_privacy.html
- [6]<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#3cf4809a6668>
- [7]<https://www.theguardian.com/technology/2013/oct/06/big-data-predictive-analytics-privacy>
- [8]https://www.huffpostbrasil.com/entry/china-surveillance-camera-big-brother_us_5a2ff4dfe4b01598ac484acc