NOM : Juillard

PRENOM : Sandrine

FILLIÈRE : Télécommunication

Why are we scared of Big Data ?

Data analytics and privacy on the Internet

Word Count : 2 499

Submission Date : December 18th

Table of content

Introduction

In the past years the growth of computing power has been exponential. To illustrate this, the Numerical Wind Tunnel, the most powerful supercomputers of the 1990s developed by the National Aerospace Laboratory of Japan and the firm Fujistsu; a device that took up an entire room; has been outdone by the Iphone 5 (Cinquin, 2014). This technical breakthrough paved the way for what we call Big Data. Indeed, BigData refers to all processing techniques dealing with a tremendous quantity of information. Treating that much data in a humanly acceptable time requires computer power (≃speed) and storage. Hence, the rapid enhancement of our systems has opened new compelling perspectives. In plenty of fields, Big Data marked a new era and in particular with the following techniques: Predictive Analytics and user Behavior Analytics. As their name suggests, they aim to predict phenomena or anticipate and influence human choices based on the analytics of previously collected data.

A common example is Netflix recommendation : based on series and films you watched and liked, the Netflix algorithm can predict what series you may like by comparing you with users that have similar taste. This algorithm, which is only one example among all those used by Netflix, aims to optimize every aspect of the platform leading the customer to spend more time on it, and be more dependent on it.

In fact, analyzes of data can be very effective to determine, predict, and influence one's behavior and beliefs. Companies and advertisers we use daily, such as Google, Facebook or Amazon, which are nowadays essential to our lives, are of course well aware of their power. Hidden inside “terms of use” we agree blindly to, we grant permission to applications, not only to collect and use our data, but also to sell them!

Recent scandals have emerged related to this “data trading”: Big Data is believed to have a link with Donald Trump's victory in the presidential Election and the Britain's Brexit (Patel, 2018). This situation raise several a question.

The first part of this report is dedicated to explaining what is big data and its usages. We will focus on analytics in marketing and in what extent advertisement with big data can be considered as manipulation.

In a second part, we will highlight threats caused by the business around big data. In particular, we will discuss privacy issues and the dangers of the effectiveness of analytics.

I - Data Usage

1- Big Data and Data analytics

The smart equipment we use daily nowadays, produces and accumulates numerical data. It was measured in 2013, that the digital world has produced more than 3 thousand trillion data (Furutamag, 2014). This mean that in two years, we double the quantity of data produced ever. These facts highlight how the use of data has been normalized. We live in a data society, where every available piece of information is collected and saved. This enormous quantity of data is what we call Big Data. All alone, these data does not mean much but all together they can be a relevant material to be analyzed. BigData application cases are very diverse. From simple uses like storing data into an easily readable database, or the creation of statistics, to more ambitious purposes such as predicting future occurrence of natural disaster, crises as petrol shortage, or armed conflicts.
Here is some example of application of big data having incident on our everyday life:

Retailers/Service providers use analytics with their customers data in order to understand their needs and buying habits. Moreover, they can predict potential trends, or give pertinent product recommendations.

Healthcare use patient data to improved diagnostics and provide treatment options. (Illustration Figure 1)

Banks & Financial institutions use analytics to predict probable loan defaulters, customer churn out rate, and detect frauds in transactions. In addition, analytics can provide advice on financial decisions, according to the current and forthcoming economical picture. (Ingram Micro ,2017).

Science collecting data about natural phenomena, measured from space or wildlife. Among all use cases, the more remarkable one is the implication of big data on climate change detection.

Figure 1 - Data analyses in Healthcare
Source : Digital Buisness and buiness analytics - Timo elliott's blog

Taking a step back, analyzing the past to predict the future and improve the present is something humankind has always been doing. The real novelty is what the digital revolution and data abundance brought to the world of analyzes. Data analyzes have been taken to a whole new level and have become standard as an essential tool in numerous fields.

2- Analytics in marketing

Communication and marketing have always tried to classify the population into groups in order to identify customers based on their gender, age, income, etc. For instance, the advertisement and packaging of a watch have to fit the taste of the target audience's average profiles. Data Analytics work on the same concept to a greater extent. Groups are way more distinctive and personalized considering the quantity of information available about the customers.

The idea is to make smaller and more specific groups based on other criteria such as opinions, interests, lifestyles. Hence, instead of sending the same message to all “40-year-old Housewives”, or “Teenager from a modest family” which is highly ineffective, message/advertisement will be address to subgroups of the audience. By collecting one's data, and comparing them to other users, analytics is able to create an accurate portrait of one's personality and thus, found the most accurate way to reach the user. “Tell me what you like, I will tell you who you are” said John Ruskin during the nineteenth century.

Netflix uses this mechanics to analyze films you watched and recommend you those you might also like. Google and Facebook use the same mechanism to suggest advertisement that might interest you.

3- Technophobia

In light of the above, Big Data and Data analytics are wonderful technologies that can brings a lots to the world. Plus, it only reproduces thinking patterns that marketing have always been using. Therefore, why are we afraid of Big Data, and why so much buzz around it? Are we all being technophobic toward algorithms than exceed us by its accuracy and effectiveness? Well, not exactly. In fact, we haven't approached the main issue of big data: Privacy.

Figure 2 - Technophobia

Source : Observation on science, fiction and futurism - Steven Lyle Jordan

II - Privacy issues

1- Lifetime of our data

Data are collected almost everywhere. From internet sources of course, like websites you subscribe to, or from the connected object you own where is stored your social media profile, your web mails, your Google search history or the locations you visited. In addition, some more confidential information is saved on private servers, as for instance records of your electricity supplier which stored customer's power consumption or the police criminal records.This information can tell us a great deal about one's life, directly or indirectly. A criminal record is undeniably a very indicative piece of information, whereas, the potential of information given by a power consumption report is more subtle but as equally reveling (Mashima, 2018).

Indeed, as explained above, data alone are unreliable, but together they can be a very accurate tool's profiling. The sum of all these information can determine gender, political and religious beliefs, sexual orientation, and preferences of all kind.

In addition, of a distressing privacy intrusion, the leakage of those highly confidential information can expose us to a variety of very serious threats such as identity thief's, hacking, or manipulation. However, this information is essential to ensure our applications will work properly. Most of the common services we use mention their data privacy policy in the well known “terms of use” we never read. In order to have the most accurate overview of threats we incur by a neglected privacy protections, they will be separated into two categories : data trading and data leakage.

2- Business around data

Donald Trump's election in 2017 was orchestrated by Russian, using big data as manipulation tool. Is this high profile scandal around big data well-founded? First, let's debunk the speculations : No evidence was found to confirm a potential interference of Russia in the Presidential Campaign in 2016 (Le Monde, 2018). Unfortunately, the same cannot be said for Facebook and Cambridge Analytica.

Cambridge Analytica is a British company which uses big data as a communication tool. Their slogan : “Data drives all we do”. Despite their alarming watchword, private companies as well as governmental organizations are counted among their customers. 6 million dollars were spent by Donald Trump during his campaign for these services and a specific department of the firm, self-described as “global election management agency” was created (Audureau, 2018). Yet, the activities and usages of Cambridge Analytica are perfectly legal and was known long before the scandal broke out. No judicial proceedings were instituted for their use of Big Data.

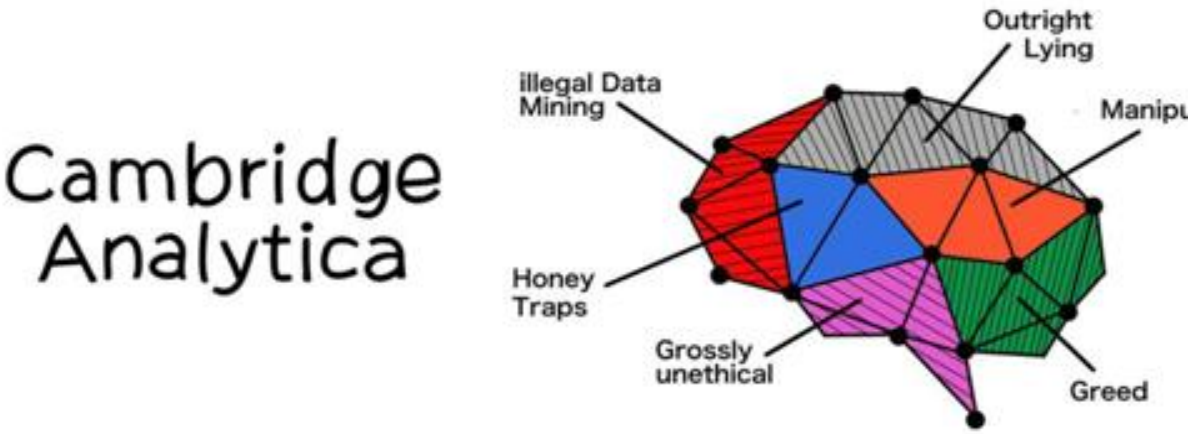

Figure 3 - How Cambridge Analytica Used Big Sleaze To Mine Big Data
Source : Forbes cartoons

The activities of Cambridge Analytica is fearsome (illustration figure 3) and its efficiency is not to prove, but despite it all, communication tools using big data is considered like any other tools of communication. In fact, as every technique that aim to influence people (advertisement, eyes-catching packaging, etc), can be considered somehow as manipulation. The issue is the gap of efficiency and scaling between traditional communication and big data based communication. Thus, those two practices are equal toward the law, regardless of their non-equivalence ethically. In addition, governmental actions to regulate this situation are very limited, if not nearly non-existent.

In fact, the doubtful source of their datasets is at the actual origin of the incrimination. Millions of Facebook users' personal data was collected by Cambridge Analytica, through an innocuous-looking app which consisted in a series of question to build psychological profiles. This application, harvested all personal data available from the Facebook profile of its user, and also those of their Facebook friends.

Indeed, algorithms of analytics developed by Cambridge Analytica is useless without proper dataset. Hence the importance of a strong and efficient privacy politics from data providers and a responsible attitude from the users. Some malicious app mention inside their “term of use” a clause that grant them permission to use/sell data's from their user. Hence, unmindful user expose themselves to this kind of threats, by accepting those conditions.

3- Data leakage

Even if we consider services providers as fully honest, risk still exist. Many services have insufficient preventing systems against cyber-attacks and spying leading content of the database to be stolen. Depending on the content of the system hacked, gravity of the damage can vary from low to extremely severe as to directly jeopardizing one's integrity; regardless of the issues we mention earlier.

The most common data to be stolen is the password. Plenty of people, despite warnings, use the same password on every application they use. On a small-scale, vulnerable applications, databases are more easily hackable. Then, when passwords are collected, malicious persons can program software in order to try all the e-mail/password set they harvested as login of other services, in particular commercial websites where payment cards could be saved. Passwords are not the only data usable against someone though. Other data can be used to create personalize fishing mails in order to intimidate and improve truthfulness of a scam, or explicitly as blackmailing material, against the person hacked and its contacts.

Being careful of one's information usage thus, is essential. Regulation laws to prevent business around data is not sufficient. These facts highlight the importance of awareness and self-responsibility of users to preserve their privacy and safety, for themselves and for their relations.

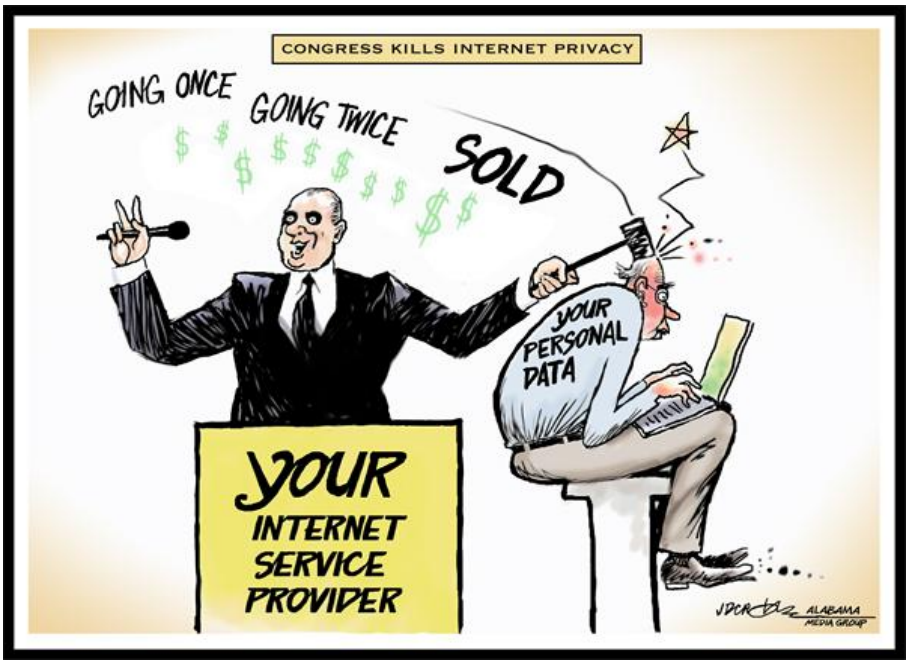

Figure 4 - Internet privacy Auction
April 2017 - Debate around enabling the Internet Services Provider (ISP) to create fast lane/slow lane depending on contract prices. In addition of receding net neutrality, it would have create issues of privacy since the ISP would collect data from users. (New York Time, 2017)
Source : Cagle - J.D. CROWE

Conclusion

Both malicious peoples and advertisers, are always looking for the most complete and detailed profiles to increase their values as material for spying and manipulation. Therefore, the normalization of firm privacy policies and the use of advanced security techniques by our applications seem to be an issue of great concern. We can imagine a future where artificial intelligence and Data analyses assist humankind, bringing out its greatest potential. The perspective of a society enhanced by its technological advancement is compromised if the security of its users is not guaranteed.

Private Web is an active area of research and new innovative solutions are published frequently. However, these solutions have a specific scope of protection and are not adapted for all kind of use. The most popular solution is Tor which prevent services to collect data from users. Although this solution is not unerring and slow down internet access. Despite being open to all kind of users, this protocol is mostly use for illegal practices as drug sells on internet, or by the opposition in authoritarian regimes. (The guardian, 2013)

Plus, it does not prevent the risk occurred by users when data are consentingly given. Hence, it cannot preclude neither social medias nor hackers to collect personal information.

There is an urgent matter about trust users can grant to internet companies. According to Neil M. Richard and Jonathan H. King, the current awareness and precautions taken are far from being enough (O'Flaherty, 2019). Furthermore, they suggest establishing ethical principles for the sake of values as free choices, transparency, individuality and, needless to say, privacy. Moreover, we have not mentioned a major key issue of Big data development: its excessive energy consumption. The ascendancy of big data is widely recognized; hence our society should assess scaled security systems, making both companies and individuals more accountable. This can only be archived if governments enforce strict laws to prevent companies from abusing their potency.

In the USA, no federal law protect personal data (ICLG, 2020) and the situation seem to evolve in the wrong direction (illustration figure 4). However, since 2016, the European union agreed on regulation applying to European and Foreign societies, restricting their rights on personal data collected. (CNIL, 2016).

Issues bringing out by the Big data have become a keenly debated topic and we should expect evolutions in line with European measures in hope of brighter prospects.

References

[1] Ce que vous avez toujours voulu savoir sur le Big Data, Ludovic Cinquin, USI, 2014

[2] Role Of Big Data In US Presidential Election, Pranay Patel, Mount Royal Univerity, 2018

[3] Les enjeux du Big Data, Furutamag, ARTE, 2014

[4] People Fear Workplace Data Because They Don't Want To Be Held Accountable, Max Nisen, Buisness Insider,2013

[5] 5 Big Data Use Cases in Banking and Financial Services, Ingram Micro ,2017

[6] What Are the Biggest Privacy Issues Associated with Big Data? Cyber security insiders, 2020

[7] How to find out what Google knows about you and limit the data it collects, Todd Haselton, CNBC, 2017

[8] Duality of Big Dat, Daniel Riedel, wired, 2018

[9] Huawei Security Schandal, Kate O'Flaherty, Forbes, 2019

[10] Big data Ethics, Neil M. Richard and Jonathan H. King, Washigton University School of Law, 2014

[11] Towards quantitative evaluation of privacy protection schemes for electricity usage data sharing, D.Mashima, A.Serikova, Y.Cheng, B.Chen, 2018

[12] Ce qu'il faut savoir sur Cambrige Analytica, William Audureau, Le Monde, 2018

[13] Le règlement général sur la protection des données ,CNIL, 2016

[14] USA: Data Protection Laws and regulation, ICLG (International Comparative Legal Guide), 2020

[15] What is Tor? A beginner's guide to the privacy tool, The Guardian, 2013

[16] FCC Repeals Net Neutrality, New York Times, 2017