Big Data Dangers: Weaponizing Social Media

Ross Wood
rmw@indiana.edu
HID 345

ABSTRACT

Social media has changed the way people get information, disseminate information, communicate, and stay in touch with others, both online and in the real world. As more and more people from different age groups and socioeconomic backgrounds begin adopting social media and becoming active users, data is being created at a geometric rate. The analysis of all this data being generated can be used in a myriad of different ways, including nefarious ones. It is possible to analyze the digital footprint of social media users in order to accurately target enormous swaths of a population with propaganda, misinformation, and deception which have been created to cater to the specific population's social or political bias.

KEYWORDS

i523, HID345, Social Media, Social Media Mining, Big Data, Social Media Scraping

1 INTRODUCTION

The rise of social media among all tiers of society, not just the tech savvy portion, has created interesting opportunities in the field of big data and social media mining. The decrease in costs and size of computing tools is also helping to fuel a technological explosion among different societies, with more and more people having access to social media than ever before. This increase in users creates a tremendous amount of data, all of which can be analyzed to reveal information about the users and real world populations. This information can be used to inform, educate, and improve the lives and systems we use daily. However, in the wrong hands, this kind of information can also be used to influence and manipulate citizens into supporting things that are against their self-interest, and against the interests of their society. If an informed citizenry is essential to maintaining freedoms, then in effect, social media can be weaponized and used to curtail freedoms for some by misinforming and radicalizing its user bases.

2 USER BASE EXPLOSION

The increase in population of social media user bases is helping lead the way in 21st century social engineering, for better or worse, and this increase is causing data to be created at a scale that has never before been seen. Indeed, a report found that the population of adults in the United States who use social media rose from 7% in 2005, to 65% in 2015 [6]. Furthermore, the report found that "there continues to be growth in social media usages among some groups that were not among the earliest adopters, including older Americans" [6]. The ability to scrape massive amounts of user data from social media sites allows for an analysis of a user's individual digital footprint. When all these footprints are put together and analyzed, conclusions about an individual's taste in entertainment, political, and social leanings can be drawn, as well as information about an individuals personality and socioeconomic background. When these conclusions are combined with user location and network structure data, a situation is created where an organization or group could manipulate an entire segment of a country's population. The success of this method depends largely on how accurate all the accumulated user data is. The more accurate the data, the better job a machine does at making these demographic predictions. But just how accurate can a machine be at predicting a human's personality and sociopolitical leanings based on digital information alone?

2.1 Accuracy

The accuracy of a machine's prediction about individual personality and demographics is improved as it accumulates more user data to work with. In essence, the more a person uses social media and creates information about themselves, the easier it is going to be for a machine to look at this information and predict certain things about the person. One study found that to a certain point, humans are better than machines at making personality judgments on other humans. However, once a machine has, in this example as little as 100 Facebook likes, the machine's ability to make accurate personality judgments starts to outperform the predictive ability of humans [9]. The study found that with even a small amount of data, machines can predict a person's personality better than that same person's close acquaintance. The study's findings also "highlight that people's personalities can be predicted automatically and without involving human social-cognitive skills" [9]. This automated process makes it easy for people and organizations to gather large amounts of user data for analysis, that can then be used however these people or organizations see fit.

3 NEFARIOUS DEMOGRAPHIC INFORMATION ANALYSIS

An individual's personality traits are not the only information that machines can gleam from peoples' social media footprints. Indeed, there are a number of different methods of analysis that can be used when approaching this problem. One would use different methods for trying to figure out different kinds of population information during analysis. It is possible to infer user data, such as age, race, and socioeconomic background, based only on concepts as simple as word use and assigning emotional feelings to different words [7]. All of this demographic information would be quite valuable to individuals or organizations wanting to understand or influence various populations and subsets of populations, or even just to understand user bases for marketing purposes. This user population information that has been generated could then be used to target these various population segments with misinformation and propaganda that appeals to their specific confirmation bias. If the misinformed user were to then share the misinformation or propaganda with their like minded friends online, this could create a self-sustaining cycle of misinformation that reinforces the

misinformed beliefs of users. This is known as an echo chamber and they can be quite pervasive in social media [2].

3.1 Using Demographic Data

There are recent examples of these techniques having been used on populations. A post electoral analysis of various elections in Russia found that not only are governments using this approach to their benefit, but also that it works best in regions that exhibit large amounts of racial, social, religious, or socioeconomic tension [4]. Using this data to spread misinformation works by causing both sides of any argument to seem radical, even the rational side. The misinformation does this by working off the confirmation bias of the reader, which was acquired by analyzing their digital footprint. The example of Russian election use found that this approach worked best in areas that were particularly volatile in regards to racial prejudice and struggle [4]. So in other words, someone with a racial bias towards a certain group would have their beliefs reinforced through social media use. This effect is only further reinforced by social media users whose experience on social media is limited primarily to echo chambers, which further distort their view of society while simultaneously widening the societal divide and radicalizing users [2].

3.2 Bots

Misinformation and propaganda, which add fuel to the fire of discussions in these so called echo chambers, can also be spread by the insidious use of bots, which are programs that do automated tasks. These tasks range from simple jobs like retweeting something, to complex tasks like conversing with a human and tricking them into thinking the bot is real. Whatever the case, that task is often one that helps convince people of an agenda that the bots have been told to push. As of March 2017, there are almost 48 million active twitter bots. These bots can push any information their masters want, while also serving the purpose of inflating the popularity of people, tweets, and points of view that are more aligned with the bots' agenda [1]. Bots are key to the successful spread and propagation of misinformation and their campaigns. They are now capable of even greater insidiousness by being able to target specific groups using user data generated by social media users. This misinformation technique is taken to another level with bots sophisticated enough to engage in limited conversations with real people over social media [8]. Chat rooms, message boards, and social media sites are flooded with bots, all pushing different agendas. The more resources and effort an organization can put behind an misinformation push of this manner, the more more effective it will be.

3.3 Social Polarization

Irregardless of who uses these techniques and for what purposes, one outcome that always arises from their use in this way is social polarization. The focal point of the polarization depends on which social struggle is being exploited to manipulate the common social media user. Protests, economic inequality, racial discrimination: there are any number of current social problems to draw from if one wanted to fan the flames of social unrest on a large scale in order to push a political or corporate agenda. This effect is beginning to have a visible influence on societies around the world as misinformation campaigns are causing more and more people to be misinformed on current events. A misinformed voter does not make good choices, and enough of them together has the potential to throw the entire democratic process out of whack [3]. Social polarization leads to instability and unrest, which can be profitable to certain members of society who might take advantage of these techniques.

4 DISCUSSION

Social media use has become so ingrained into everyday life that, at this point, it would be almost impossible to get people to stop using it, even if it was demonstrated to them that it has a potential negative effect on society. Indeed, this abstinence approach should not be advocated, as social media and the data it produces can be used to benefit society at large in a multitude of ways. That being said, it would be wise to continue to monitor and study different ways social media can harm society by being used to benefit the few at the expense of the many [5]. If processes and protections aren't put into place in the near future, the entire democratic process could continue to destabilize and become polarized to the point where it is so consumed by corruption that it cannot be salvaged.

One possible solution is to examine your network and establish who the key peddlers of misinformation are and to block or delete their accounts. This approach would be effective since "successful sources of false and biased claims are heavily supported by social bots" [8]. Another approach is to fight fire with fire and create bots that are sophisticated enough to detect misinformation as it begins to trend and then counter this trend with the truth [1]. Whatever solution to this problem takes shape, the exponential growth of social media users and the unprotected data they generate [6] make it imperative that a solution is found and implemented. Until that happens, huge portions of social media users are going to continue to be tricked into believing misinformation and propaganda through data analysis and manipulation.

5 CONCLUSION

The rising population of social media users is beginning to pose a threat in regards to a population's ability to stay accurately informed. As this population of users grows and creates more and more data, so to does the ability to use sophisticated techniques to deceive the users. The growth of social media grows hand in hand with new dangers. As more people get their news through social media, it becomes easier to misinform them. In essence, the more information that is known about someone, the easier it is to take advantage of them. And if you are trying to dupe someone, the modern world makes it easy to accumulate information on people by analyzing their social media digital footprint.

We are starting to see real world effects of these techniques in the form of population destabilization and user manipulation through propaganda and misinformation campaigns. At present there are no safeguards in place to protect users from attempts to deceive them, no matter where the attacks come from or what agenda they have. Until safeguards are developed and put into place to protect users and the enormous amounts of data that they generate from those who would use it against them, the problem is only going to continue to grow and get worse.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski, Miao Jiang, and Juliette Zerick for assistance with this assignment and using github.

REFERENCES

[1] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the Spread of Misinformation in Social Networks. In *Proceedings of the 20th International Conference on World Wide Web (WWW '11)*. ACM, New York, NY, USA, 665–674. https://doi.org/10.1145/1963405.1963499

[2] Siying Du and Steve Gregory. 2017. The Echo Chamber Effect in Twitter: does community polarization increase?, Vol. 693. 373–378.

[3] Robert Epstein. 2016. *Subtle New Forms of Internet Influence Are Putting Democracy at Risk Worldwide.* Springer New York, New York, NY, 253–259. https://doi.org/10.1007/978-1-4939-6415-4_9

[4] Regina Goodnow, Robert Moser, and Tony Smith. 2014. Ethnicity and Electoral Manipulation in Russia. 36 (12 2014).

[5] Rodrigo Ochigame and James Holston. 2016. Filtering Dissent Social Media and Land Struggles in Brazil. *New Left Review* 99 (Jan. 2016), 85 – 100. https://newleftreview.org/II/99/rodrigo-ochigame-james-holston-filtering-dissent

[6] A. Perrin. 2015. *Social Media Usage: 2005-2015: 65% of Adults Now Use Social Networking Sites–a Nearly Tenfold Jump in the Past Decade.* https://books.google.com/books?id=OupAnQAACAAJ

[7] Daniel Preoţiuc-Pietro, Svitlana Volkova, Vasileios Lampos, Yoram Bachrach, and Nikolaos Aletras. 2015. Studying User Income through Language, Behaviour and Affect in Social Media. *PLOS ONE* 10, 9 (sep 2015), e0138717. https://doi.org/10.1371/journal.pone.0138717

[8] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. 2017. The spread of fake news by social bots. (07 2017).

[9] Wu Youyou, Michal Kosinski, and David Stillwell. 2015. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences* 112, 4 (jan 2015), 1036–1040. https://doi.org/10.1073/pnas.1418680112

[6] [9] [7] [4] [2] [1] [3] [5] [8]

6 BIBTEX ISSUES

Warning–numpages field, but no articleno or eid field, in Budak2011

Warning–empty booktitle in Du2017

Warning–empty publisher in Du2017

Warning–empty address in Du2017

Warning–can't use both author and editor fields in Epstein2016

Warning–no journal in Goodnow2014

Warning–page numbers missing in both pages and numpages fields in Goodnow2014

Warning–empty publisher in Perrin2015

Warning–empty address in Perrin2015

Warning–no journal in Shao2017

Warning–no number and no volume in Shao2017

Warning–page numbers missing in both pages and numpages fields in Shao2017

(There were 12 warnings)

7 ISSUES

Have you written the report in the speciffied format? - citation "blocks" are listed sequentially at the end of the references section, probably just an artifact; beautifully written paper