

A critical analysis of big data for international development

Sven Preng

May 25 2020

1 Introduction

The usage of big data has already shown to be a valuable asset for companies that wish to innovate their products and/or services based on the growing amount of data. While these innovations have mainly been focused on products for urban areas in the global north, researchers have been investigating breaking ground with usage of big data in rural areas in the global south. Analysis of big data could be beneficial for international development and to achieve the goals for sustainable development, introduced by the United Nations (UN). However, the term big data has been around for 15 years at the time of writing and during this time, some pitfalls became clear for both the global paradigm and, more specific, big data to aid development. This essay will focus on the usage of big data in current development work and research and how the pitfalls could affect international development

2 The characteristics of big data

The characteristics of big data usage have been explored by Hilbert [1]. During the last 20 years, traceable information transaction has skyrocketed both in developing and developed countries. Search queries, user behavior and preferences, to name a few, are a by-product of the digitization that has been going on. The capacity to store data also increased as hard disks became more affordable. However, during this period of development and growth in valuable data, the capacity to analyze this data started lacking behind. While computation and storage capacity did improve over this time, as predicted by Gordon Moore in 1965 [2], the sheer amount of produced data introduced complexities of its own. When working with these large amounts of data gets complex for even the most advanced technologies, it can be classified as big data [3]. Companies and educational institutions worked on various approaches to address these complexities. This led to the introduction of MapReduce[4], the initial successful big data model, in various incarnations. While MapReduce is a model that aims to make the analysis of large amounts of data manageable to a certain

degree of accuracy, it is not a silver bullet. Various successors developed after MapReduce, solve some of its problems but introduce new challenges. Since its introduction big data has been an ever-changing landscape of technologies and its evolution is far from over. For companies with the available funds, adapting is not a big problem but developing countries could suffer from the growing pains of big data if the required experience is unavailable.

3 D4D

Companies and researches involved in developing countries have been working with big data analysis for the last decade. An example discussed by Taylor and Schroeder is Orange's D4D challenge [5]. During this UN endorsed project aimed at Ivory Coast and Senegal, Orange offered anonymized data to researchers in an effort to address development problems. This project is notable, as it was the first project, aimed at development and thus received an endorsement by the UN. As this project was first of its kind, it showed some of the obstacles the various involved parties must overcome. Firstly, governments like the Ivory Coast government, may or may not have access to citizen data and thus cannot verify estimations done by research teams. Secondly, companies like Orange must think about political risks of sharing their data because large data sets like this offer many possibilities which could be of aid but could also be exploited. And finally, international development workers must coordinate the efforts of the other involved parties to ensure the local context is understood correctly. While the findings did provide some use for data scientists, they have yet to be useful for Ivory Coast and Senegal as no policy adjustment has been made.

4 Flowminder

Not all efforts of data scientists to analyze data are meaningless for the development country at hand. Flowminder was a project that was initiated after the earthquake at Léogâne, Haiti in 2010. [5] About 10 months after the earthquake, the first cholera outbreak in recent history took place in rural areas of the country after the destruction the country had suffered. Flowminder was established to predict movement of the population in order to restrain the spread of the disease. In order to do so, phone companies gave the Flowminder team access to SIM card information. This information allowed the team to localize people and follow their movement in a time span from a month before the earthquake until several months after the earthquake. While this project was successful, it introduced some major problems. The first one being the possible bias towards the user base of mobile phones in Haiti. Several vulnerable groups like children, the elderly and some of the poorest people of the country, could have been excluded from the findings as their data was not included. While this was not a major issue in the case of this particular disaster, other similar efforts could suffer from comparable biases. The second issue is the privacy of the user

that is at stake. While major disaster response usually condones the usage of privacy sensitive information, not all crises are similar. An example could be the 2020 COVID-19 pandemic in the Netherlands during which citizens showed concern about their privacy in the possible usage of a mobile app which could help with containment of the virus. [6]

5 Conclusion

Big data is still a relatively young phenomenon in the field of data analysis. While it does serve a purpose to major players in the global north, its fitness for the global south needs to be contemplated. Firstly, working with large amounts of data is difficult, even for researchers that have the the most advanced technologies available. Adapting these technologies in the global south might be demanding due to the lack of experts in the area. Furthermore, Orange's D4D project shows that while companies may have their best intentions for international development, and external data analysts offer their knowledge to help, it is still difficult to work on actual policies in response to the research. However, not all big data for development projects have been in vain. The Flowminder project shows that crisis response is a great target for working with big data. Nevertheless it is important to keep the privacy sensitivity of the data at hand in mind. It is of top priority that experts in the field of big data have a good grasp of the local context of their efforts to be of real value. Assisting local experts with new technologies could also make sure that developing countries actually get the help that is required.

References

- [1] M. Hilbert, "Big data for development: A review of promises and challenges," *Development Policy Review*, vol. 34, no. 1, pp. 135–174, 2016.
- [2] G. E. Moore *et al.*, "Cramming more components onto integrated circuits," 1965.
- [3] A. Jacobs, "The pathologies of big data," *Queue*, vol. 7, no. 6, pp. 10–19, 2009.
- [4] T. Condie, N. Conway, P. Alvaro, J. M. Hellerstein, K. Elmeleegy, and R. Sears, "Mapreduce online." in *Nsdi*, vol. 10, no. 4, 2010, p. 20.
- [5] L. Taylor and R. Schroeder, "Is bigger better? the emergence of big data as a tool for international development policy," *GeoJournal*, vol. 80, no. 4, pp. 503–518, 2015.
- [6] NOS, "Privacywaakhond: 'burger geen wandelende antenne in contact met overheid'," Arpil 7 2020. [Online]. Available: <https://nos.nl/collectie/13824/artikel/2329754-apps-moeten-verspreiding-coronavirus-tegengaan-maar-hoe-zit-het-met-privacy>