

Big Data Analytics in Developing Countries

Judy Phillips
Indiana University
Bloomington, Indiana
judkphil@iu.edu

ABSTRACT

Developing nations cope with numerous humanitarian challenges. Infrastructures are often inadequate to deal with basic public health, public safety, and environmental concerns. As a result, citizens deal with issues such as poverty, food insecurity, and the unavailability of basic health care. Resource limitations often make it difficult to manage crisis situations such as natural disasters. The use of wireless and Internet related technology is growing globally. Mobile phone and social media usage are becoming common even in remote areas. As a result, Big Data analytics is playing a role in mitigating the impacts of some of these humanitarian concerns.

KEYWORDS

I523, HID332, developing countries, food insecurity, public safety, big data

ACM Reference format:

Judy Phillips. 2017. Big Data Analytics in Developing Countries. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17)*, 4 pages.

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

Individuals in developing nations face a long list of humanitarian challenges, including poverty, hunger, health care access, and availability of clean water sources. Other challenges include insufficient resources to deal with public safety and crisis intervention issues.

The statistics are dismal. "Almost 1.3 billion people living in developing countries live on less than 1.50 dollars a day" [2]. "According to the United Nations, approximately twenty two thousand children die each day in these countries due to poverty" [1]. More than eight hundred seventy million people in third world nations have no food to eat or a very precarious food supply. "A third of all childhood deaths in sub-Saharan Africa is caused by hunger related diseases" [1]. That is approximately 2.6 million deaths per year. One child dies every five seconds of starvation [1]. More than two hundred million children under five years of age in developing countries do not reach their developmental potential due to malnutrition [7]. Over 1.2 billion people around the globe do not have regular access to clean drinking water. Many people die from

common curable diseases that such as malaria, pneumonia, and diarrhea because they do not have access to health care. Approximately ten million children die each year from treatable diseases. [2]. Fifty percent of pregnant women in developing countries lack proper prenatal care. This results in over three hundred thousand maternal deaths annually from childbirth. [1]. The threat of HIV is also reaching a pandemic level in many of the third world countries [1].

Big Data Analytics is starting to be used to address some of these issues. Digital data is becoming more widely available globally. Internet wireless communications and mobile phone access are starting to become commonplace even in some rural areas. The data collected from these devices is being combined with data collected via traditional data sources such as datasets and surveys. This is providing information and insights that have never before been available. "The diffusion of data science into the realm of international development constitutes an opportunity to bring powerful new tools in the fight against poverty, hunger, and disease" [5]. Furthermore, the real time availability of much this data enables more timely and agile implementations of solutions. This all results in significantly better outcomes.

2 INFRASTRUTURE

In recent years, there has been a huge increase in the availability of digital technology globally, including in developing nations. According to the GSM Association 79 percent of the worlds total inhabited areas had mobile network coverage in 2012 [4]. According to the International Telecommunications Union, there were 2 billion people using the internet in 2015 and there were 91.8 mobile phone subscriptions per 100 inhabitants in developing countries [4]. Social media such as Facebook and twitter is being utilized by more and more people worldwide. Sensor technology is becoming less expensive and more efficient. Better algorithms are being developed to utilize the lower cost sensors for developmental activities. Data information sources include call logs, mobile banking transactions, blog posts, tweets, and Facebook content [5].

The diffusion of mobile phone technology has been especially important. Because mobile phones are often the only interactive technology that low income individuals have access to, they have become the cornerstone of many Big Data projects in the developing world [4].

3 BIG DATA

The amount of data that is being generated in developed countries is increasing rapidly. According to the Cisco Global Cloud index the highest workload growth rates between 2013 and 2018 are expected to be in the Asian Pacific, the Middle East and Africa, and Latin America. Growth rates during these time periods are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2017 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

expected to be 45 percent, 39 percent, and 34 percent respectively. "Data center traffic in the Middle East and Africa is expected to reach 366 exabytes in 2018 compared to 68 exabytes in 2013" [4].

4 HEALTHCARE

"Big Data has enormous potential to address health care challenges in the developing world" [4]. One of the primary problems with healthcare in the developing world is the overall lack of access. This is caused by a combination of geographical accessibility and the lack of basic medical resources. There are shortages trained medical professionals, medical equipment, and drug stocks. People in rural areas often have to travel long distances in order to obtain care. There are also a lack of resources to implement basic public health regimes such as immunization policies. All of this makes the occurrence of serious disease outbreaks and epidemics common and difficult to manage when they do occur. Another issue is the existence of widespread fake drug distribution networks.

4.1 Public Health

One area in which Big Data can have an enormous impact on the health of vulnerable populations is in public health policy. Proper public health infrastructure is needed to prevent, treat, and manage serious disease outbreaks. Public health policies and related public education can also educate populations and influence attitudes and behaviors concerning important health related matters such as maternal health and immunizations.

Big Data is extremely useful for managing serious disease outbreaks, including pandemics and epidemics. Big Data and data science can be used first to track and monitor the spread of the disease and then to effectively allocate resources and medication so that the disease can be properly treated and contained. In fact, the term for this field is Infodemiology. It is a whole new field of data science [5].

Health related data is mined from social media and sites such as twitter and then combined with data visualization techniques to track the geographic spread of a disease. As the spread of the disease is being tracked in real time, big data is used to ensure that all available resources are allocated effectively. Big data ensures the right distribution of resources, including medical personnel and medication at the right time to the right location [6]. Proper resource allocation is especially important when lifesaving medical supplies are in short supply. According to the US Center for Disease Control and prevention (CDC), online data can help detect disease outbreaks before confirmed diagnosis or lab confirmation [5]. It is estimated that disease outbreaks can be identified up to two weeks sooner than with the use of traditional methods such as physician reporting [4]. When this resource allocation technique was used in Tanzania during a malaria outbreak, it reduced the number of drug facilities that were out of stock of the appropriate medication during the epidemic from 78 percent to 26 percent [4].

Social Media can also be used to track peoples health related beliefs, perceptions and concerns at any a given time and in real time. This methodology is referred to as sentiment analysis. For example, researchers can get an indication of health related attitudes about immunizations, the use of medication or prenatal care programs by reviewing social media posts. These studies can assist with health

related education efforts. Social media and big data analytics are also be used to measure the impacts of humanitarian aid and intervention. For example, the United Nations used this technique to evaluate whether the Every Woman Every Child initiative had had an impact. This was a program that was designed to increase awareness of maternal health, breastfeeding, vaccinations. A team of researchers analyzed social media posts for two years for relevant keywords, such as breastfeeding or vaccination to determine if the program has resulted in increased parental awareness [4]. The information collected can be used to identify needs in order to establish and manage public health policies and programs.

Sentiment analysis can also be used to track other public health related issues such housing shortages, employment, and inflated food prices. This methodology is able to identify issues earlier than traditional methods and thus enables more timely deployment of resources and solutions [5].

4.2 Health Care Access

In developing countries, there are often problems with geographical accessibility to health care. People in rural areas often need to travel long distances to visit a health care professional. Also, rural areas do not have enough primary health care providers and specialists are rarely available.

The Internet of Things technology can solve some of these issues. One solution is patient sensors. Relatively low cost sensors can be worn on the person to monitor physiological variables in real time. The data collected can be transmitted to health care providers in a distant locations for diagnosis and treatment. These sensors can be used for routine as well as critical health issues such as heart palpitations. For example, in Africa there is a device called Cardio pad. It is a medical tablet that can be used to perform and collect information from cardiology related tests by individuals who have no cardiac training. The information gathered can then be sent to a cardiac specialist via mobile phone in order to receive diagnosis and treatment instructions[4]. In China the Internet of Things technology Institute is developing a telephone booth sized health capsule. Rural villagers can be receive a diagnosis from a distantly located physician when they step into it. [4].

4.3 Distribution of Fake Drugs

The widespread distribution of fake drugs is a huge health hazard in developing nations. According to the World Health Organization, counterfeit antimalarial and tuberculosis drugs account for seven hundred thousand deaths annually. Big Data technology is playing a huge role in fighting this crime. One nonprofit organization has developed a possible solution. The name of the program is called GoldKeys. All legitimate prescription containers have a twelve digit scratch off code. Customers can verify the authenticity of the medication by texting the scratched off code number to a health hotline. The number is matched to information in a cloud database and the information is sent back to the customer. The project is being maintained and funded primarily by Hewlett Packard [4].

5 ENVIRONMENTAL PROTECTION AND WATER SUPPLY

Almost a billion people in the world do not have a reliable source of clean drinking water [1]. According to World Water Development Report in 2012, inadequate sanitation and poor hygiene result in 3.5 million deaths annually [4]. Much of the water is wasted or leaked due to faulty pipes. Other water is lost due to unidentified or unnecessary pollutants.

The Internet of Things can be used for the purpose of monitoring water supply and quality. Sensors are frequently used to monitor pollutants in a river or water source. Resources are deployed to remedy problems when they are detected. One example is in the city of Da Nang, Vietnam. Da Nang is a major port city on the South China Sea. The Da Nang water company uses Big Data to provide real time analysis of the city's water supply. The goal is to better manage leaks, monitor pollutants, and accurately forecast future demand. Big Data sensors are installed throughout each stage of the water treatment process. Water quality is tracked in real time. Notifications are sent if there are problems. [4].

In another example, IBM worked with the city of Tshwane in South Africa to develop a crowd source application that users use to report water supply issues such as faulty pipes. The result was the discovery of thirty million dollars of wasted water sources. This application operates without the need of a central inspection authority [3].

6 PUBLIC SAFETY AND CRISIS INTERVENTION

One of the most important areas in which Big Data is being deployed is to enhance public safety and crisis intervention efforts during natural disasters. "The availability of digital data collected and analyzed rapidly and in real time can drastically improve interventions and outcomes in crisis situations for vulnerable populations" [5].

One of the most widely used tools in this effort are crisis maps. Crisis maps use data from numerous sources, including local citizen reports, social network data, and environmental data to aid emergency responders in times of natural disaster. "Crisis maps have been deployed during dozens of events worldwide, including the 2012 Haiti earthquake and the 2010 Pakistan floods" [3]. In Haiti during an earthquake a centralized text message center was set up that allowed cell phone users to report where people were trapped. The United States Geological Survey has developed a system that monitors twitter for spikes about earthquakes globally. This information can be used to evaluate the location, quantify magnitude, identify epicenter, and respond quickly and appropriately [5].

7 AGRICULTURE

"More than half the population in all of the developing nations depend upon agriculture and farming for at least two meals a day. This accounts for almost seventy five percent of the world's poorest people" [1]. Therefore, one important way to address poverty and food insecurity is to find ways to make farming techniques more effective and productive. Big Data has big potential to dramatically increase production for small scale farmers.

"Studies suggest that ineffective farm operations such as late planting, lack of proper land preparation, improper harvesting

techniques and poor housing and feeding of livestock can reduce a smallholders farmers productivity by up to forty percent" [4]. One technique for improving production is Precision agriculture. The objective of Precision agriculture is to provide farmers with informed, personalized information so that they can make better operational decisions in real time. Data is collected on things such as soil conditions, weather, seeding rates, and crop yields using technology such as sensors, drones and satellites [4]. Sensors can be located in fields, inside livestock, or on farm equipment. After the data is collected it is analyzed and returned to the farmers via computers and mobile phones in terms of customized solutions. Instructions may be such things as the optimum type of seeds, pesticides, herbicides, and fertilizer use. The objective is to match inputs with the exact need. When resources are used efficiently production is maximized. Another solution involves collecting data to locate and notify farmers of the spread of crop and livestock plights. The objective is that farmers take safety measures as soon as possible [3].

In Uganda there is a Big Data tools project that uses Precision agriculture techniques that were developed by the Grameen Foundation. Data is collected on farmers, farming practices, and external conditions. It is given back to farmers in the form of a community knowledge database via Android phones. Information about the time and methods of planting crops, caring for farm animals and marketing their products [4].

Another way in which big data can be used for small holder farmers to support financing opportunities. In Nairobi, Africa the company Gro Ventures is building a platform which integrates information about crops and the environmental conditions to give lenders more confidence to lend money to farmers. One of the offerings allows farmers to pool their data to apply for collective loans to buy shared tractors and equipment [3].

8 CHALLENGES

There are many challenges to the successful implementation of many of these projects. Many people in the least developed nations still lack access to internet service or a mobile phone. There are high costs associated with using big data technology. Cost of mobile phones, analytical services and data services often cost prohibitive for individual citizens. There is also a Big Data skill set deficient. Big data technology and the analytics to turn big data into actionable information requires technical skills that are often not available. Furthermore, health care professionals and other related personnel often lack knowledge or training about data science.

In order for initiatives to be successful, financial and technical support will need to come from other sources: academia, public and private sector, and philanthropic. To date, there are numerous non-government organizations (NGOs) working throughout the world to fight poverty and reduce disease [6]. The United Nations started an initiative in 2009 called Global Pulse. The objective of Global pulse is to research ways that Big Data can be incorporated into the developing world to improve lives. They are currently conducting several research initiatives in various locations throughout the world. Several private organizations are also playing a role. For example, Google has announced a plan to develop high speed internet solutions in developing countries using high altitude balloons.

Their goal is to add an additional 1 billion people to the Internet from Africa, and Southwest Asia [4].

9 CONCLUSION

Although Big Data does not have the ability to solve all of the world's problems, it does have enormous potential to reduce suffering and save lives for those living in developing countries. Big data is giving smallholder farmers resources to substantially increase their food production. This will play a substantial role in the fight against poverty and food insecurity. Big data analytics is improving health by making health care accessible to even those in the most remote locations. Big data provides the knowledge to identify and monitor water availability issues such as waste and pollution so that problems can be identified and dealt with immediately. Big is also saving lives by providing the real time knowledge needed to respond effectively to health epidemics and natural disasters. As the use of internet related devices continues to increase throughout the developing world, the impact of big data will continue to grow.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski and the teaching assistants in the Data Science department at Indiana University for their support and suggestions to write this paper.

REFERENCES

- [1] ELIST10. 2014. Top 10 Major Problems in Third World Countries. Web page as Article. (June 2014). <http://www.elist10.com/top-10-major-problems-third-world-countries/>
- [2] Institute of Ecolonomics. 2015. Top 5 Challenges the Third World is Facing Today. Web page as Blog. (May 2015). <http://ecolonomics.org/top-5-challenges-the-third-world-is-facing-today/>
- [3] Travis Korte. 2014. How Data Analytics Can Help the Developing World. Web page as Article. (Sept. 2014). https://www.huffingtonpost.com/travis-korte/how-data-and-analytics-ca_b_5609411.html
- [4] Nir Kshetri. 2016. *Big Data's Big Potential in Developing Economies*. CABI.
- [5] United Nations Global Pulse. 2012. Big Data for Development Challenges and Opportunities. Web page as paper. (May 2012). <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGloalPulseMay2012.pdf>
- [6] Mark Van Rijmenam. 2017. How Big Data Analytics Can Help the Developing World Beat Poverty. Web page as Article. (July 2017). <https://datafloq.com/read/big-data-developing-world-beat-poverty/168>
- [7] Wikipedia. 2017. Developing Country. Web page. (Oct. 2017). https://en.wikipedia.org/wiki/Developing_country