Applications and Practices of Big Data for Development

Harshini Priya Adusumalli[1*], Mahesh Babu Pasupuleti[2]

[1]Department of Computer Science, Kent State University, Kent, Ohio, **USA**
[2]Data Analyst, Department of IT, Nitya Software Soluctions INC, Fremont, CA 94538, **USA**

[*]E-mail for correspondence: harshinipa.gs@gmail.com

https://doi.org/10.18034/abr.v7i3.597

ABSTRACT

Daily, vast volumes of public data are generated due to the expansion of social media sites, digital computing devices, and Internet connectivity. Effective data analysis techniques/algorithms can provide near real-time knowledge regarding emerging patterns and early warning in case of an emergency (such as the outbreak of a viral disease). These data can also disclose several helpful indicators of socioeconomic and political events that can assist formulate effective public policy. This study examines the use of big data analytics for human development. As big data technology matures, it will be possible to use it for development purposes, such as addressing humanitarian crises or violent conflicts. Large-scale use of big data for development is fraught with obstacles due to its huge quantity, rapid change, and diversity. The most urgent challenges are effective data collection and exchange, providing context (e.g., geolocation and time), and ensuring data accuracy and privacy. This study reviews existing big data for development studies to assess the impact of big data on society's development. We examine major efforts while also highlighting obstacles and unresolved issues.

Key words: Application Development, Human Development, Big Data, Big Data Analytics

INTRODUCTION

Big data, or the tremendous increase in data, is both an opportunity and a threat to researchers. Processing, storing, and analyzing large data has come a long way: In addition to big data computing (processing and storing big data on a cluster of computers), rapid advances in intelligent data analytics (AI and ML) provide the ability to process massive amounts of diverse unstructured data generated daily to extract valuable actionable knowledge. This allows academics to use the data to develop meaningful knowledge and insights.

Companies like Google and Facebook deal with petabytes of data. Google handles about 24 petabytes of data per day, while Facebook receives over 10 million photographs per hour. We are inundated with data as a result of rapid technological advancements (e.g., the Internet of Things (IoT), which uses sensors, for example in the form of wearable devices, to offer data relating to human activities and diverse behavioral patterns).

A key challenge for big data for development is acquiring access to crucial people-related data, which is often only available to governments in the form of paper papers. Fortunately, the new "open data" trend promotes open public sharing of data from various public and commercial sector organisations in searchable and machine-readable formats (Pasupuleti, 2015a). The US and UK governments are increasingly adopting open data projects to foster innovation and openness. Also, open source platforms have been developed to collect digital data from mobile devices (e.g., the Open Data Kit). While open data is a subset of all available big data, the nuance is in big data's liquidity. Various hackathons are being organized to tap the potential of open data in terms of useful mobile applications (e.g., the local government of Rio de Janeiro has created the Rio Operation Center aimed at harnessing the power of technology and big data to run the city effectively in terms of transport management, natural disaster relief, mass movement and management of slum areas). According to a recent McKinsey Global Institute analysis, open data is worth $3 trillion. Less open data means less transparency, which means more transparency for consumers.

BACKGROUND: BIG DATA TECHNIQUES

Modern datasets, also known as big data, differ from old datasets in three ways: they are larger in number, faster in processing speed, and more diverse in content. In today's world, enormous volumes of data are being generated at an astronomical rate (or velocity), and the different sources of data provide a plethora of options for analysis.

It is possible to realize the concept of the information age if all of this data is intelligently harnessed and utilized. The data can be transformed into actionable information by applying clever processing and analytics to the data that has already been collected. This section is concerned with the techniques (particularly those connected to machine learning) that are used to collect, store, process, and evaluate the massive amounts of data that are generated. We also make an attempt to connect this conversation, as well as the many examples used to clarify various principles, to the field of humanitarian development (Pasupuleti, 2015b). While the purpose of this part is to give readers with a brief background on the relevant methodologies and associated work in order to assist them comprehend their applicability when presented from the perspective of humanitarian development, it also serves another purpose.

Numeric prediction

Rather of forecasting the discrete class (or category) to which the example belongs, we are interested in predicting the numeric quantity associated with the example in numerical prediction. Think about the weather dataset that was previously mentioned to describe association learning as an example once again.

Data mining

Data mining is a term that refers to the automated finding and prediction of patterns in vast amounts of data using machine learning techniques (Witten and Frank, 2005). Data mining can also refer to online analytical processing (OLAP) or SQL queries, which require searching a huge database for a certain query after it has been collected in a previous period of time.

Machine learning

In artificial intelligence (AI), machine learning (ML) is a sub-field that is concerned with the issue of enabling computational systems to learn from data about how to accomplish a desired task automatically. Machine learning has a wide range of applications, including decision-making, forecasting, and predicting, and it is a key enabling technology in the deployment of data mining and big data techniques in a wide range of fields, including healthcare, science, engineering, business, and finance. Machine learning is also becoming increasingly popular in the financial sector.

Supervised learning

It is the job of ML algorithms in this class of learning to generalize from a training set, which is labeled by a "supervisor" to contain information about the class of an example, in order to make predictions about new examples that have not yet been seen. Regression is used to describe problems in which the output (or prediction) is a continuous set of values, whereas classification is used to describe problems in which the output (or prediction) is a discrete set of values (or prediction).

Unsupervised learning techniques

Unsupervised learning relies on clustering. In clustering, the objective is to categorize examples into 'clusters' based on perceived similarity. This clustering is used to locate comparable input groups. While supervised classification requires a correctly labeled training set, unsupervised clustering aims to directly discover the structure of input data.

Reinforcement learning

This is a machine learning technique that is based on rewards and punishments. This strategy involves a learner who, in response to an input received, takes some action that may have an impact on the environment around him or her. Following that, the action is either rewarded or punished. Probabilistic mapping describes the relationship between the actions done by the learner and the rewards or penalties that result from those activities. The ultimate goal of a learner is to identify such an ideal mapping (or policy), from its actions to the rewards and penalties, in order to maximize the average long-term reward.

Deep learning

Deep learning (DL) is a machine learning technique that use deep and complex architectures to learn from data (Schmidhuber, 2015). Each of these processing layers is capable of generating a non-linear response that corresponds to the data input. These architectures have numerous processing layers. These layers are made up of a number of tiny processors that work in parallel to process the data that is delivered. Neurons are the processors that make up this system. DL has proven to be effective in a variety of applications, including pattern identification, image processing, and natural language processing.

New trend in database technology: NoSQL

The introduction of big data and Web 2.0 has resulted in the creation of a massive volume of unstructured data such as word documents and emails as well as blog posts, social networking content, and multimedia content. It differs from structured data in that it cannot be organized and saved in the usual relational databases, but structured data can be organized and stored in relational databases. Unstructured data must be stored and accessed using a different strategy and set of procedures than structured data. NoSQL (or non-relational) databases were created for the same purpose as relational databases (Leavitt, 2010).

Predictive analytics

Data mining and machine learning techniques are used to predict future events or behavior. Predictive analytics is a

technology that tries to create a competitive advantage by forecasting some future events or behavior (using historical data and machine learning techniques) (in the form of collected data). Predictive analytics is a broad term that includes data science, machine learning, predictive and statistical modeling, and it produces empirical predictions based on empirical data that is provided as input (Shmueli and Koppius, 2010).

Crowdsourcing and big data

Crowdsourcing differs from traditional outsourcing. The difference between crowdsourcing and traditional outsourcing is that a task or a job is outsourced, but not to a specific professional or company, but to the whole public through an open call. Crowdsourcing is a strategy that can be used to acquire information from a variety of sources, including text messages, social media updates, blogs, and other online forums (Pasupuleti, 2015c).

Internet of things

The Internet of Things (IoT) is a new hot sector that has been propelled by the hype surrounding big data, the advent of network science, the expansion of digital communication devices, and the availability of ubiquitous Internet connection to the general public (Adusumalli, 2016a). According to a technical analysis published by the McKinsey Global Institute, the Internet of Things has the potential to provide significant economic value.

BIG DATA FOR DEVELOPMENT: DEVELOPMENT AREAS

In this section, we will discuss some of the important development areas where big data can be used for development purposes. We will first look at the importance of big data in development during natural disasters and political upheavals, and then we will go on to other topics. We are also investigating how Big Data for Development might be applied in the domains of agriculture, healthcare, education, and the alleviation of poverty and hunger, in addition to these humanitarian crises (Adusumalli, 2016b).

Humanitarian emergencies: This study will give two case studies involving natural catastrophes and political crises, through which we will demonstrate the significant role that big data may play in a variety of situations. Different challenges relating to the collecting, storage, and exchange of data in the event of an emergency are also taken into consideration.

Hunger, food and agriculture: Kshetri (2014) conducts an examination of contemporary research literature as well as government reports/documents in order to determine the elements that facilitate the use of big data techniques for development goals, as well as the ones that stand in the way of this process. The value of new data sources, such as social media and cell phone data, is emphasized in this article. Nonuniform spread of technological advances and trends throughout the world is a result of differences in skill levels, financial ability to afford data, and, in some cases, cultural and industrial standards for exploiting current technological innovations. With the help of a case study in agriculture, this paper will address the opportunities and obstacles associated with the application of big data techniques for the development of farmers' businesses (Adusumalli, 2017a).

Healthcare: Healthcare organizations are undergoing a significant culture shift as they embrace big data analytics to improve diagnosis and treatment outcomes for their patients. Incorporating data from a patient's various medical records, as well as data from real-time wearable sensors, can revolutionize medical diagnosis by allowing doctors to analyze and diagnose the patient's current health status, as well as provide an early warning sign if the patient's health is on a dangerous track. This aids in the collection of preventative measurements that can be used to diagnose and treat a potentially hazardous disease in its early stages.

Education: The field of education is undergoing a shift to a digital age, with the usage of traditional textbooks dwindling and the use of digital versions of study materials becoming increasingly popular. Education is one of the industries that has reaped the benefits of big data analytics to a significant degree. Revolutionary shifts are taking place in international pedagogical techniques, students' learning and study habits, and the overall design and operation of the educational system (Adusumalli, 2017b).

BIG DATA ANALYTICS FOR DEVELOPMENT

Mobile analytics

It is the application of big data approaches to huge amounts of data that mobile businesses collect on their users in terms of call volume, calling pattern, and location. Mobile analytics is a subset of data science that is growing in popularity. It contains a variety of information that can be extremely beneficial for study, planning, and development (the use of such information also poses many privacy and ethical use challenges). Specifically, the field of mobile big data analytics is concerned with analyzing cell-phone data in order to generate insights that may be used to develop value-added services. It is possible to extract socioeconomic information from mobile service providers' "call-detail-records" (CDR) analysis, which is held by the companies that supply the services. A project by Nokia Research called the Mobile Data Challenge (Laurila et al., 2012) was one of the initiatives targeted at collecting and utilizing mobile phone data for research purposes. The paper (Laurila et al., 2012) provides a detailed description of the study, its objective, and the research methods. It is estimated that about 200 smart-phone users in Switzerland provided their mobile phone data for the purpose of this research project. For a variety of developmental reasons, such as urban planning and transportation engineering, analysis of social

dynamics of a group of people, and even epidemic control, mobile analytics can be applied.

Living analytics

Living analytics is an interdisciplinary research topic that combines skills from computer science, network science, social science, and statistics. Living analytics is the study of individual and group social and behavioral tendencies. The discipline of computational social science is built on exploiting developments in storage and computing capabilities to process publicly available big data to advance our understanding of social science.

Visual analytics

Visual analytics is a fascinating subset of big data exploration that aims to enable analytical thinking using visual interfaces. Large amounts of quantitative data can be seen in a small area. This section will include data maps, time series, space-time narratives, and relational graphics. We summarize Tufte's (1983) approaches and relate them to humanitarian development.

Time series

This style of graph shows a variable's growth, development, decay, or general trend through time. Time resolutions range from seconds to centuries. Time series include the rise and fall of the stock market, regional temperature variations, and user device usage patterns (Eagle and Pentland, 2006). Time series are vital for analyzing trends throughout time (e.g., dengue mosquitoes that mostly bite during dawn and dusk and during specific months of a year). This type of information allows disaster management authorities to take proactive measures to reduce casualties.

CHALLENGES, OPEN ISSUES AND FUTURE WORK

Given the current state of technology, it is highly likely that big data will gain significant significance and potential in order to transform the paradigm of the conventional humanitarian development process in practically every field of endeavor. It is not, on the other hand, a panacea for all of the problems that face society today (Pasupuleti, 2016b). There are numerous possible hurdles to large-scale implementations of big data, just as there are for any other breakthrough. Some of these problems are discussed in this section from two different perspectives: the technological and ethical. In response, we discuss open issues and the work that will be required to overcome these obstacles in the coming months.

Technical challenges

In order to use big data for development, there are a number of technical hurdles to overcome. What about the processing and storing capacities, for example, are they increasing in tandem with the massive volumes of data being produced on a daily basis? Following are a few of the technological issues that we have encountered and described:

- When we explored the migrant problem, we noted the relevance of crowdsourcing. Facebook, Twitter, and other social media platforms are excellent sources of crowdsourced data, and many relief organizations rely on the information gathered from these sources.
- Personalized material predicted by algorithms based on a user's past behavior can lead to polarization when presented to that user. This means that two separate users could receive completely different search results when searching for the same thing. The use of new deep-learning approaches, which do not totally rely on previous data, and the use of context aware computing and algorithms can help to alleviate these problems.
- Despite the numerous advantages of using big data for policy analysis, it is a dangerous undertaking. This procedure is fraught with numerous possible difficulties and dangers. When collecting information from users, privacy is a fundamental topic that has been hotly disputed. The context and semantics of data might be affected during the data collection process (the big-data supply chain), resulting in erroneous and often controversial policies.
- A large number of people update their statuses with information on a catastrophe while all congregating at a single geographical location. This presents a hurdle in terms of identifying the real location of the crisis for which the information was provided in the first place. Consequently, the data acquired from actual ground surveys and aerial imaging should be validated with these in order to take effective action in a crisis situation when one occurs.
- Large-scale data analytics frequently involve the collection and subsequent combining of unstructured data from a variety of data sources.
- The problem of fragmentation is one of the most significant impediments to the widespread adoption of big data analytics on a large scale. To give an example, a patient may be seen by several different doctors for what appear to be disparate medical issues.
- Cloud computing and software-defined networking (SDN) technologies have proven extremely useful for efficiently implementing big data solutions in recent years; however, more work will be required in the future to ensure that computing and networking infrastructure can scale to accommodate the ever-increasing volume of data.

Ethical challenges

In addition to all of the technology-related problems discussed above, it is critical to consider the ethical dimensions of using big data for development purposes. Over the course of the article, we have attempted to explain, in addition to all the benefits, the potential problems and downsides that could result from the deployment of Big

data for development purposes (Pasupuleti, 2016a). We discovered that privacy is one of the most significant concerns in practically every industry where big data analytics is used. In addition to privacy concerns, the difficulty of fragmentation is a significant hurdle to the widespread adoption of big data analytics at a broad scale. Aside from these well-known difficulties, there are a few more subtle ones to contend with, the majority of which fall under the headings of ethics and technological exploitation. The advancement of data science is a difficult task in and of itself. As a result, competence and collaboration among people from a variety of subjects and disciplines are required in the sector. To evaluate big data with the appropriate views and ethics in place, inter-disciplinary initiatives should be encouraged and financially incentivized, as well as financially rewarded.

CONCLUSIONS

Despite its enormous potential, big data for development research is still in its infancy. Because development is a multifaceted subject, we choose to approach the problem of big data for development from a multidisciplinary perspective (integrating technology, economics, and social and development sciences). We reviewed existing research, official documents, online projects, blogs, and technical publications connected to big data for development for this work. Our research outlines some of the underlying problems and potential hidden harms that must be addressed and countered. Unlike other survey articles, we examine numerous methodologies for big data analytics, as well as outstanding concerns and future research objectives. We reviewed the literature on leveraging big data for human development in this research (big data for development). Our goal is to raise awareness about the enormous potential of big data for development in areas such as humanitarian catastrophes (disaster response and migration crisis), agriculture, poverty reduction, food production, healthcare, and education. We have discussed the risks and disadvantages of using big data for development. We expect big data for development to play an important role in future human and global development, but researchers must be able to address and overcome the problems outlined.

REFERENCES

Adusumalli, H. P. (2016a). Digitization in Production: A Timely Opportunity. *Engineering International*, *4*(2), 73-78. https://doi.org/10.18034/ei.v4i2.595

Adusumalli, H. P. (2016b). How Big Data is Driving Digital Transformation?. *ABC Journal of Advanced Research*, *5*(2), 131-138. https://doi.org/10.18034/abcjar.v5i2.616

Adusumalli, H. P. (2017a). Mobile Application Development through Design-based Investigation. *International Journal of Reciprocal Symmetry and Physical Sciences*, *4*, 14–19. Retrieved from https://upright.pub/index.php/ijrsps/article/view/58

Adusumalli, H. P. (2017b). Software Application Development to Backing the Legitimacy of Digital Annals: Use of the Diplomatic Archives. *ABC Journal of Advanced Research*, *6*(2), 121-126. https://doi.org/10.18034/abcjar.v6i2.618

Eagle, N., Pentland, A. (2006). Reality mining: sensing complex social systems. *Pers Ubiquitous Comput*, *10*(4), 255–68.

Kshetri, N. (2014). The emerging role of big data in key development issues: Opportunities, challenges, and concerns. *Big Data & Society*, *1*(2), 2053951714564227.

Laurila, J. K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T-M-T., Dousse, O., Eberle, J., Miettinen, M. (2012). The mobile data challenge: Big data for mobile computing research. In: Proceedings of the Workshop on the Nokia Mobile Data Challenge, in Conjunction with the 10th International Conference on Pervasive Computing, p. 1–8.

Leavitt, N. (2010). Will nosql databases live up to their promise?. *Computer*, *43*(2), 12–4.

Pasupuleti, M. B. (2015a). Data Science: The Sexiest Job in this Century. *International Journal of Reciprocal Symmetry and Physical Sciences*, *2*, 8–11. Retrieved from https://upright.pub/index.php/ijrsps/article/view/56

Pasupuleti, M. B. (2015b). Problems from the Past, Problems from the Future, and Data Science Solutions. *ABC Journal of Advanced Research*, *4*(2), 153-160. https://doi.org/10.18034/abcjar.v4i2.614

Pasupuleti, M. B. (2015c). Stimulating Statistics in the Epoch of Data-Driven Innovations and Data Science. *Asian Journal of Applied Science and Engineering*, *4*, 251–254. Retrieved from https://upright.pub/index.php/ajase/article/view/55

Pasupuleti, M. B. (2016a). The Use of Big Data Analytics in Medical Applications. *Malaysian Journal of Medical and Biological Research*, *3*(2), 111-116. https://doi.org/10.18034/mjmbr.v3i2.615

Pasupuleti, M. B. (2016b). Data Scientist Careers: Applied Orientation for the Beginners. *Global Disclosure of Economics and Business*, *5*(2), 125-132. https://doi.org/10.18034/gdeb.v5i2.617

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Netw*, *61*, 85–117.

Shmueli, G., Koppius, O. (2010). Predictive analytics in information systems research: Robert H. Smith School Research Paper No. RHS, 06–138.

Tufte, E. R. (1983). The Visual Display of Quantitative Information. CT: Graphics press Cheshire. Graves-Morris P, Vol. 2.

Witten, I. H., Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques: *Morgan Kaufmann.*

--0--

Online Archive: https://abc.us.org/ojs/index.php/abr/issue/archive