**Uncertainty in big data analytics: opportunities, challenges and survey**

Authors: Het Shah, Kashyap Shah, Meet Shah

**Abstract**

Big data analytics has received widespread attention from both academics and industry as the need to understand trends in large databases has increased. Recent advances in sensory networks, cyber-physical systems, and Internet of Things (IoT) have increased data collection (including health care, communications, smart cities, agriculture, finance, education, and more) on a much larger scale. However, data collected by sensors, social media, financial records, etc. Uncertain due to noise, imperfection, and inconsistencies. Analysis of such large amounts of data requires advanced analytical strategies for effective review and / or predicting future studies with high accuracy and advanced decision-making strategies. As the quantity, variability, and speed of data increase, so does internal uncertainty, leading to uncertainty in the analysis process and decisions made as a result. Compared to data strategies and traditional platforms, artificial intelligence techniques (including machine learning, natural language processing, and computer intelligence) provide accurate, fast, and awesome results for large data analytics. Previous research and research conducted on big data analytics usually focuses on one or two processes or specific domains of the application. However, very little work has been done in the field of uncertainty when used in large data analytics and in artificial technology methods used in data sets.

**Keywords**

* Data
* Big data
* Big data analytics
* Data analysis

**Introduction**

According to the National Security Agency, the Internet processes 1826 petabytes (PB) of data per day. In 2018, the number of data generated per day was 2.5 quintillion bytes. Previously, the International Data Corporation (IDC) estimated that the amount of data produced would double every 2 years, but 90% of all data in the world was produced in the last 2 years, and moreover Google is now considering more than 40,000 searches every second or 3.5 billion searches per day. Facebook users upload 300 million photos, 510,000 comments, and 293,000 status updates per day. Needless to say, the amount of information produced on a daily basis is staggering. For this reason, strategies are needed to analyze and understand this large amount of data, as it is a good source from which you can find useful information.

Advanced data analysis techniques can be used to convert big data into intelligent data for the purpose of obtaining sensitive data in relation to big data. Therefore, intelligent data provides useful information and improves decision-making capacity in organizations and companies. For example, in the healthcare sector, analytics performed on large data sets (provided by programs such as Electronic Health Records and Clinical Decision Systems) could allow health care workers to deliver effective and cost-effective solutions to patients by examining the patient's general history, compared to reliance on evidence provided with localized data or current data. Big data analysis is difficult to perform using traditional data analytics as they may lose performance due to the five V characteristics of big data: high volume, low veracity, high velocity, high variability, and high value. In addition, many other features are present in larger details, such as variability, viscosity, accuracy, and performance. Many artificial intelligence (AI) techniques, such as machine learning (ML), natural language processing (NLP), computational intelligence (CI), and data mining were designed to provide great data analysis solutions as they can be faster, more accurate, and more accurate. in addition to large amounts of data. The purpose of these advanced analytical methods is to obtain data, hidden patterns, and anonymous integration into large databases. For example, a detailed analysis of a patient's historical data may lead to the detection of devastating diseases at an early age, thus enabling appropriate treatment or treatment program. In addition, risky business decisions (e.g., entering a new market or starting a new product) can benefit from better imitation of decision-making skills.

While large amounts of data use AI hold many promises, a variety of challenges are presented when such strategies are faced with uncertainty. For example, each V element presents multiple sources of uncertainty, such as random, incomplete, or noisy data. In addition, uncertainty can be embedded in the entire mathematical process (e.g., collecting, editing and analyzing big data). For example, dealing with incomplete and ambiguous information is a critical challenge for many ML data mining methods and techniques. In addition, the ML algorithm may not receive the best results if the training data are biased in any way. Wang et al. presented six key challenges in the analysis of big data, including uncertainty. They focus more on how uncertainty affects learning performance from big data, and a separate concern lies in reducing the uncertainty that exists on big data. These challenges often arise from data mining and ML strategies. Measuring these concerns to a large extent will data effectively cover any errors or shortcomings of the entire mathematical process. Therefore, reducing uncertainty in big data analysis should be at the forefront of any automated process, as uncertainty can have a significant impact on the strength of its results.

Based on our current research findings, little has been done about how uncertainty significantly affects the combination of big data and analytical techniques used. To address these shortcomings, this document presents an overview of existing AI strategies for the analysis of big data, including ML, NLP, and CI in view of the uncertain challenges, as well as the appropriate guidelines for future research in these domains. Contributions for this work are as follows. First, we look at the challenges of uncertainty in each 5 V data signal. Second, we review a few strategies in large data analytics that contribute to the uncertainty of each method, and we also review the impact of uncertainty on many data analysis strategies. Third, we discuss the strategies available to deal with each challenge presented by uncertainty.

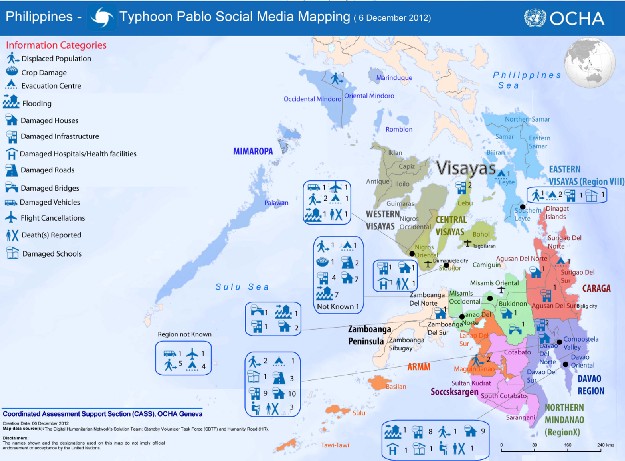
**Will be increase in big data analysis Job Opportunities**

As organizations begin to realize that they cannot use big data in the process of photographing, interpreting and using that data, they have begun to look for professionals who can do so. Just look at any major job site and you will find that there are a lot of job postings by companies looking for data analysts. This number will eventually continue to increase as the data will increase significantly and the number of professionals with the skills required for this job will remain low. So, now is the time to prepare to be part of the great big data that is yet to come.

**Role of Big Data Analysis during human crisis**

Big data can help create a clearer picture of the regional consequences of a disaster. Reports from eyewitness testimony by the Witnesses following the 2010 earthquake in Haiti indicated where the victims were buried under the rubble and in areas where help was most needed.

Following Typhoon in the Philippines, more than 20,000 communications messages were used by the Digital Humanitarian Initiative to create a map of the impact of the hurricane and to determine the immediate area of ​​help.



This map is the first ever official UN crisis map entirely based on data collected from social media.

According to a report compiled by CNN, the U.S. Department of State has analyzed the data to prevent conflict from escalating. Behavioral and ethical trends in social media are analyzed by the Conflict and Stabilization Operation office to identify threats in order to put in place the necessary strategies to prevent outbreaks of violence.

As a data analytics professional, you’ll have a wide range of job titles as well as domains from which you can choose according to your preference. Since data analytics is used in different fields, lots of job titles like big data engineer, big data analytics architect, big data analyst, big data solution architect, analytics associate, big data analytics business consultant, metrics and analytics specialist etc will be available to you. Also, an array of top organizations like Microsoft, IBM, Oracle, ITrend, Opera are utilizing big data analytics and thus huge job opportunities with them are possible.

**Future priority for big companies and organisations**

New technologies in the field make it possible to perform complex data analysis tasks across various and larger datasets. Many experts use advanced data analysis techniques and tools to perform tasks such as data mining, forecasting statistics, among others. With the huge data analytics that provide businesses with a competitive edge, companies use a variety of analytical tools on an ongoing basis. Today, it is almost impossible to find a high-quality product that does not take advantage of at least some form of data analysis. Due to the growing rate of data acquisition analytics, it can be said that the future of big data analytics will hold a good position for skilled professionals.

In order to keep business competitive, top companies are looking to use data analytics to explore new market opportunities for their products and services. Today, a large percentage of large companies view data analytics as an important part of their business operations and a key way to rise above the competition and this will be even more important as competition increases over time. It means that those who wish to study today's big data will be able to be a natural part of the big data of the future.

**Big data analysis challenges**

Data security is a major problem today, with common topics related to data breach of sensitive data from large (and small) organizations. Software and security tools are already a major factor in many CIO insurance budgets. Digital data distributed worldwide connected will make that task much more difficult, requiring significant deployment of automation technology and AI. Insurance providers need to be aware of not only the potential for data theft, but also the potential for new data transactions. For example, images of property or motor vehicle damage can be altered to show serious injury or damage where there is none.

We are in the early stages of a major debate over who owns the data, who has the rights to use it, and how the data can be used. Major global technology executives who have built their businesses on data are highly regarded for these problems. Insurance providers have an additional consideration of industry regulation that sets out how data can be used in the industry. Government and industrial regulations will continue to fluctuate, which in turn creates a challenge for insurers.

Of course, there are other challenges such as finding and training talented people who can tackle these three areas and use AI technology to make sense of every detail. Both areas are important and are the subject of much research and debate, but they focus on the basic issues of data management, protection and control.

**References**

* <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#4146a89b60ba>
* <http://www.cimaglobal.com/Pages-that-we-will-need-to-bring-back/velocity-archive/Student-e-magazine/Velocity-December-2015/P2-using-big-data-to-reduce-uncertainty-in-decision-making/>
* <https://becominghuman.ai/will-you-be-a-part-of-future-big-data-analytics-3d845f2b5815>
* <https://medium.com/@JohnSnowLabs/how-big-data-and-data-analysis-can-help-in-humanitarian-crisis-mapping-312f18ef05a8>
* <https://inmediatesg.medium.com/the-challenges-in-insurance-towards-big-data-f80f22a8450c>