

BIG DATA LEADS TO EFFICIENT DECISION MAKING: IMPROVING ACCURACY

K.Krishnaveni & P.Pabitha Muthu

Assistant Professors, Departement of Computer Science Engineerereng, Pandian Saraswathi Yadav Engineering college, Sivagangai, Tamilnadu , India

Abstract

A Key part of Big Data Applications is to collect, maintain and analyze enormous amount of data efficiently. Current system, however are not able to provide accurate result. Hence it is proposed to improve the Accuracy in the Big Data Analysis. Accuracy is the foundation on which all data-based decision rest, if accuracy is not given the attention it deserves, the entire analytical and data-driven process becomes suspect. Accuracy of the result is directly propositional to the Input data and processing Algorithm. Data Quality is mainly manifested in its accuracy. Big Data also challenges the principles of Accuracy. So if possibly improve the accuracy of input and algorithm, it will result in optimized solution.

Keywords—Big Data Analytics, Accuracy, Outliers;

Introduction

Big Data has been one of the current and future research frontiers. In this year, Gartner listed the “Top 10 Strategic Technology Trends For 2013” [1] and “Top 10 Critical Tech Trends for the Next Five Years” [2], and Big Data is listed in the both two. It is right to say that Big Data will revolutionize many fields, including business, the scientific research, public administration, and so on. For the definition of the Big Data, there are various different explanations from 3Vs to 4Vs. DougLaney used volume, velocity and variety, known as 3Vs [3], to characterize the concept of Big Data. The term volume is the size of the data set, velocity indicates the speed of data in and out, and variety describes the range of data types and sources. Sometimes, people extend another V according to their special requirements. The fourth V can be value, variability, or virtual [4]. More commonly, Big Data is a collection of very huge data sets with a great diversity of types so that it becomes difficult to process by using state-of-the-art data processing approaches or traditional data processing platforms. In 2012, Gartner information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization”. More generally, a data set can be called Big Data if it is formidable to perform capture, duration, analysis and visualization on it at the current technologies.

Importance of big data

Big data is different from the data being stored in traditional warehouses. The data stored there first needs to be cleansed, documented and even trusted. Moreover it should fit the basic structure of that warehouse to be stored but this is not the case with Big data it not only handles the data being stored in traditional warehouses but also the data not suitable to be stored in those warehouses. Thus there comes the point of access to mountains of data and better business strategies and decisions as analysis of more data is always better

A. Log Storage in IT Industries

IT industries store large amount of data as Logs to deal with the problems which seem to be occurring rarely in order to solve them. But the storage of this data is done for few weeks or so though these logs need to be stored for longer duration because of their value. The Traditional Systems are not able to handle these logs because of their volume, raw and semi structured nature. Moreover these logs go on changing with the s/w and H/w updates occurring. Big data analytics not only does analysis on the whole /large data available to pinpoint the point of failures but also would increase the longevity of the log storage.

B.Sensor Data

Massive amount of sensor data is also a big challenge for Big data. All the industries at present dealing with this large amount of data make use of small portion of it for analysis because of the lack of the storage infrastructure and the analysis techniques. Moreover sensor data is characterized by both data in motion and data at rest. Thus safety, profit and efficiency all require large amount of data to be analyzed for better business insights.

C. Risk Analysis

It becomes important for financial institutions to model data in order to calculate the risk so that it falls under their acceptable thresholds. A lot amount of data is potentially Underutilized and should be integrated within the model to determine the risk patterns more accurately.

D. Social Media

The most use of Big data is for the social media and customer sentiments. Keeping an eye on what the customers are saying about their products helps business organizations to get a kind of customer feedback. This feedback is then used to modify decisions and get more value out of their business.

Big data challenges

Big data may hold a lot of potential, but it can still be held back if the data being analyzed is inaccurate. Due to restrictions on technology and other business considerations, the analyses companies are getting back may not reflect what is really happening. If businesses want to ensure their big data insights get the desired results, they need to improve the accuracy in their analytics efforts. In a perfect world, organizations would gather a vast amount of data, analyze it, and generate solutions to the problems they're facing. The truth is, as most know, we do not live in a perfect world. Insights from big data often have to be derived in a short amount of time. The technology a business has on hand might not be advanced enough to process so much information quickly. These restrictions lead many companies to performing big data analytics using sampling. In other words, they don't look at all of the data, but rather analyze only smaller subsamples of information. While this might be a go-to strategy for many businesses, the results have a greater chance of being inaccurate. Since it is vital for organizations to build accurate big data models, only looking at part of the data could lead to businesses forming the wrong conclusions.

A. Public health

analyzing disease patterns and tracking disease outbreaks and transmission to improve public health surveillance and speed response; 2) faster development of more accurately targeted vaccines, e.g., choosing the annual influenza strains; and, 3) turning large amounts of data into actionable information that can be used to identify needs, provide services, and predict and prevent crises, especially for the benefit of populations.

B. Social Media

Social media can provide instant news faster than traditional news outlets or sources and can be a great wealth of information, but there is also an ever increasing need to verify and determine accuracy of this information. Social Media Application is the best platform for understanding the real-time customer choice or intentions and sentiments. By using such data, Social media business advertising and product marketing becomes easier.

C. IOT:

IoT is not only an important source of big data, but also one of the main markets of big data applications. Because of the high variety of objects, the applications of IoT also evolve endlessly. Logistic enterprises may have profoundly experienced with the application of IoT big data. The IoT is a substantial concept of new technology; it has great potential to support an ageing society, to improve the energy efficiency and to optimize all kinds of mobility and transport.

IV. Conclusion

This Paper describes improving the accuracy of big data analysis is simply analyzing as much data as possible. If a business is only looking at a part of the dataset, they will more likely miss some of the smaller details. Perhaps a dataset contains several outliers and smaller clusters that would normally be missed. While those outliers by their very definition don't represent the whole, they may give a clue as to future trends and patterns that a company may be able to seize upon. More thorough analysis can also reveal rare events that would otherwise go unnoticed along with missing values that may indicate there is further data out there that needs to be gathered. Sampling basically decreases the accuracy of big data analysis, which limits an organization's ability to make an informed decision.

References

- [1] Eric Savitz, Gartner: Top 10 Strategic Technology Trends for 2013, October 2012.
<<http://www.forbes.com/sites/ericsavitz/2012/10/23/gartner-top-10-strategic-technology-trends-for-2013/>>.
- [2] Eric Savitz, Gartner: 10 Critical Tech Trends for the Next Five Years, October 2012.
<<http://www.forbes.com/sites/ericsavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five-years/>>.
- [3] Doug Laney, 3d Data management: controlling data volume, velocity and variety, Appl. Delivery Strategies Meta Group (949) (2001).

- [4] Paul Zikopoulos, Chris Eaton, Paul. Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Professional, 2011.
- [5] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," Commun. ACM, vol. 51, no. 1, pp. 107–113, January 2008.