IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review Study on Big Data Analysis Using R Studio on COVID 19

Praveen Kumar[1], Jyoti Kataria[2]
[1]*MTech Scholaar* [2]*Asstt. Professor*
[1,2]*Department of Computer Science & Engineering MITM, Jevra Hisar(Haryana)*
[1]mor93555@gmail.com

Abstract:- **This is focuses on scalable big-data systems, which include a set of tools and mechanisms to load, extract, and improve disparate data while leveraging the massively parallel processing power to perform complex transformations and analysis. In the near future, however, big data access at high transmission rates will be. This is a review based on COVID 19 patients data on accessible big-data systems that include a set of tools and technique to load, extract, and improve dissimilar data while leveraging the immensely parallel processing power to perform complex transformations and analysis. "Big-Data" system faces a series of technical challenges.**

Keywords: - Big Data

1 INTRODUCTION

It is obvious that we are living a data deluge era, evidenced by the sheer volume of data from a variety of sources and its growing rate of generation. For instance, an IDC report [10] predicts that, from 2005 to 2020, the global data volume will grow by a factor of 300, from 130 Exabyte's to 40,000 Exabyte's, representing a double growth every two years. The huge potential associated with big-data has led to an emerging research field that has quickly attracted tremendous interest from diverse sectors, for example, industry, government and research community. The broad interest is first exemplified by coverage on both industrial reports [2] and public media; Government has also played a major role in creating new programs [8] to accelerate the progress of tackling the big data challenges. Finally, Nature and Science Magazines have published special issues to discuss the big-data phenomenon and its challenges, expanding its impact beyond technological domains. As a result, this growing interest in big-data from diverse domains demands a clear and intuitive understanding of its definition, evolutionary history, building technologies and potential challenges.

This is focuses on scalable big-data systems, which include a set of tools and mechanisms to load, extract, and improve disparate data while leveraging the massively parallel processing power to perform complex transformations and analysis. Uniqueness of big-data, designing a scalable big-data system faces a series of technical challenges, including: First, due to the variety of disparate data sources and the sheer volume, it is difficult to collect and integrate data with scalability from distributed locations. For instance, more than 175 million tweets containing text, image, video, social relationship are generated by millions of accounts distributed globally [9].

Second, big data systems need to store and manage the gathered massive and heterogeneous datasets, while provide function and performance guarantee, in terms of fast retrieval, scalability, and privacy protection. For example, Facebook needs to store, access, and analyze over 30 petabytes of user generate data [9].

Third, big data analytics must effectively mine massive datasets at different levels in real time or near real time - including modeling, visualization, prediction, and optimization - such that inherent promises can be revealed to improve decision making and acquire further advantages.

A BRIEF HISTORY OF BIG DATA

The history of "Big Data" is presented in terms of the data size of interest. Under this framework, the history of "Big Data" is tied closely to the capability of efficiently storing and managing larger datasets, with size boundaries expanding by orders of degree.

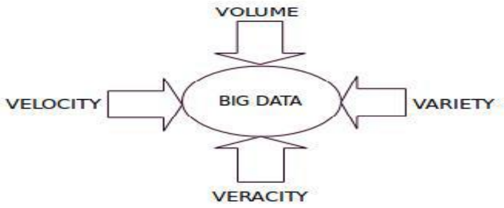

Figure 1 4V's of Big Data.

BIG DATA ANALYSIS

Big data analytics is the process of using analysis algorithms running on powerful supporting platforms to uncover potentials concealed in big data, such as hidden patterns or unknown correlations. Considering the growth and intricacy of "Big Data" science systems, previous descriptions are based on a one-sided view point, such as chronology or milepost technologies.

Descriptive Analytics: exploits historical data to describe what occurred in past. For instance, a regression technique may be used to find simple trends in the datasets, visualization presents data in a meaningful fashion, and data modeling is used to collect, store and cut the data in an efficient way. Descriptive analytics is typically associated with business intelligence or visibility systems [2].

Predictive Analytics: focuses on predicting future probabilities and trends. For example, predictive modeling uses statistical techniques [6] such as linear and logistic regression to understand trends and predict future out-comes, and data mining extracts patterns to provide insight and forecasts [4].

Prescriptive Analytics: addresses decision making and efficiency. For example, simulation is used to analyze complex systems to gain insight into system performance and identify issues and optimization techniques are used to find best solutions under given constraints.

BIG DATA PROBLEM AND CHALLENGES

However, considering variety of data sets in "Big Data" problems, it is still a big challenge for us to purpose efficient representation, access, and analysis of shapeless or semi-structured data in the further researches [12]. How can the data be preprocessed in order to improve the quality of data and analysis results before we begin data analysis [1] [2]? As the sizes of dataset are often very large, sometimes several gigabytes or more, and their origin from varied sources, current real-world databases are pitilessly susceptible to inconsistent, incomplete, and noisy data.

Therefore, a number of data preprocessing techniques, including data cleaning [11], data integration, data transformation and date reduction, can be applied to remove noise and correct irregularities. Different challenges arise in each sub-process when it comes to data-driven applications.

BIG DATA OPPORTUNITIES

The bonds between "Big Data" and knowledge hidden in it are highly crucial in all areas of national priority. This initiative will also lay the groundwork for complementary "Big Data" activities, such as "Big Data" sub structure projects, platforms development, and techniques in settling complex, data-driven problems in sciences and engineering. Researchers, policy and decision makers have to recognize the potential of harnessing "Big Data" to uncover the next wave of growth in their fields. There are many advantages in business section that can be obtained through harnessing "Big Data" increasing operational efficiency, informing strategic direction, developing better customer service, identifying and developing new products and services, identifying new customers and markets, etc.

2 R STUDIO

The R language is well established as the language for doing statistics, data analysis, data-mining algorithm development, stock trading, credit risk scoring, market basket analysis and all [9] manner of predictive analytics. However, given the deluge of data that must be processed and analyzed today, many organizations have been reticent about deploying R beyond research into production applications.

R is a statistical software, and an object-oriented high-level programming language used for data analysis, which includes a large number of statistical procedures such as t-test, chi-square test, standard linear models, instrumental variables estimation, local regression polynomials, etc. Besides, R provides high-level graphics capabilities. R is an object-oriented programming language. This means that everything what is done with R can be saved as an object. Every object has a class.

Data mining is a set of techniques and methods relating to the extraction of knowledge from large amounts of data (through automatic or semi-automatic methods) and further scientific, industrial or operational use of that knowledge. Data mining is closely related to the statistics as an applied mathematical discipline with an analysis of data that could be defined as the extraction of useful information from data. The only difference between the two disciplines is that data mining is a new discipline that is related to significant or large data sets. R is an object-oriented programming language. This means that everything what is done with R can be saved as an object. Every object has a class.

It describes what the object contains and what each function does. Application of R as a programming language and statistical software is much more than a supplement to Stata, SAS, and SPSS. Although it is more difficult to learn, the biggest advantage of R is its free-of-charge feature and the wealth of specialized application packages and libraries for a huge number of statistical, mathematical and other methods. R is a simple, but very powerful data mining and statistical data processing tool and once "discovered", it provides users with an entirely new, rich and powerful tool applicable in almost every field of research.

3 COVID 19

In the past decades, several new diseases have emerged in new geographical areas, with pathogens including Ebola, Zika, Nipah, and coronaviruses (CoVs). Recently, a new type of viral infection has emerged in Wuhan City, China, and initial genomic sequencing data of this virus does not match with previously sequenced CoVs, suggesting a novel CoV strain (2019-nCoV), which has now been termed as severe acute respiratory syndrome CoV-2 (SARS-CoV-2). Although Coronavirus disease 2019 (COVID-19) is suspected to originate from an animal host (zoonotic origin) followed by human-to-human transmission, the possibility of other routes such as food-borne transmission should not be ruled out. Coronaviruses are large group of viruses that cause illness in humans and animals. Rarely, animal coronaviruses can evolve and infect people and then spread between people such as has been seen with MERS and SARS. The outbreak of Novel coronavirus disease (COVID-19) was initially noticed in a seafood market in Wuhan city in Hubei Province of China in mid-December, 2019, has now spread to 214 countries/territories/areas worldwide. WHO (under International Health Regulations) has declared this outbreak as a "Public Health Emergency of International Concern" (PHEIC) on 30thJanuary 2020. WHO subsequently declared COVID-19 a pandemic on 11th March, 2020. Members of the family Corona virus cause a broad spectrum of animal and human diseases. Uniquely, replication of the RNA genome proceeds through the generation of a nested set of viral mRNA molecules. Human coronavirus (HCoV) infection causes respiratory diseases with mild to severe outcomes. In the last 15 years, we have witnessed the emergence of two zoonotic, highly pathogenic HCoVs: severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV). Replication of HCoV is regulated by a diversity of host factors and induces drastic alterations in cellular structure and physiology. In this review all (as we possible) information about Corona viruses are given. KEYWORDS: Corona, respiratory, viruses, Hcov, host, RNA.

SCOPE

The guidelines are in addition to the guidelines on appropriate management of suspect/confirmed case of COVID-19 issued by MoHFW on 7th April, 2020. As per existing guidelines, during the containment phase the patients should be clinically assigned as very mild/mild, moderate or severe and accordingly admitted to (i) COVID Care Center, (ii) Dedicated COVID Health Center or (iii) Dedicated COVID Hospital respectively. Guidelines for home isolation of very mild/pre-symptomatic patients were issued on 27th April 2020. The present guidelines are in supersession of the guidelines issued on 27th April 2020.

4 WHO

World health organization is providing guidance on early investigations, which are critical in an outbreak of a new virus. The data collected from the protocols can be used to refine recommendations for surveillance and case definitions, to characterize the key epidemiological transmission features of COVID-19, help understand spread, severity, spectrum of disease, impact on the community and to inform operational models for implementation of countermeasures such as case isolation, contact tracing and isolation. Several protocols are available here. One such protocol is for the investigation of early COVID-19 cases and contacts (the "First Few X (FFX) Cases and contact investigation protocol for 2019-novel coronavirus (2019-nCoV) infection"). The protocol is designed to gain an early understanding of the key clinical, epidemiological and virological characteristics of the first cases of COVID-19 infection detected in any individual country, to inform the development and updating of public health guidance to manage cases and reduce the potential spread and impact of infection.

Protect yourself and others from COVID-19. There is currently no vaccine to protect against COVID-19. The best way to protect yourself is to avoid being exposed to the virus that causes COVID-19. Stay home as much as possible and avoid close contact with others. Wear a cloth face covering that covers your nose and mouth in public settings. Clean and disinfect frequently touched surfaces. Wash your hands often with soap and water for at least 20 seconds, or use an alcohol based hand sanitizer that contains at least 60% alcohol. Practice social distancing Buy groceries and medicine, go to the doctor, and complete banking activities online when possible. If you must go in person, stay at least 6 feet away from others and disinfect items you must touch. Get deliveries and takeout, and limit in-person contact as much as possible. Prevent the spread of COVID-19 if you are sick Stay home if you are sick, except to get medical care. Avoid public transportation, ride-sharing, or taxis. Separate yourself from other people and pets in your home. There is no specific treatment for COVID-19, but you can seek medical care to help relieve your symptoms. If you need medical attention, call ahead. Know your risk for severe illness Everyone is at risk of getting COVID-19. Older adults and people of any age who have serious underlying medical conditions may be at higher risk for more severe illness.

Recommended Test

Real time or Conventional RT-PCR test is recommended for diagnosis. SARS-CoV-2 antibody tests are not recommended for diagnosis of current infection with COVID-19. Dual infections with other respiratory infections (viral, bacterial and fungal) have been found in COVID-19 patients. Depending on local epidemiology and clinical symptoms, test for other potential etiologies (e.g. Influenza, other respiratory viruses, malaria, dengue fever, typhoid fever) as appropriate. For COVID-19 patients with severe disease, also collect blood cultures, ideally prior to initiation of antimicrobial therapy.

Management of COVID-19

In the containment phase, patients with suspected or confirmed mild COVID-19 are being isolated to break the chain of transmission. Patients with mild disease may present to primary care/outpatient department, or detected during community outreach activities, such as home visits or by telemedicine. Mild cases can be managed at Covid Care Centre, First Referral Units (FRUs), Community Health Centre (CHC), sub-district and district hospitals or at home subject to conditions stipulated in the home isolation guidelines available at Detailed clinical history is taken including that of co-morbidities. Patient is followed up daily for temperature, vitals and Oxygen saturation (SpO2).

Counsel patients with mild COVID-19 about signs and symptoms of complications that should prompt urgent care. Patients with risk factors for severe illness should be monitored closely, given the possible risk of deterioration. If they develop any worsening symptoms (such as light headedness, difficulty breathing, chest pain, dehydration, etc.), they should be immediately admitted to a Dedicated Covid Health Centre or Dedicated Covid Hospital.

5 Conclusion

Big data analytics is the process of using analysis algorithms running on powerful supporting platforms to uncover potentials concealed in big data. It is obvious that we are living a data deluge era, evidenced by the sheer volume of data from a variety of sources and its growing rate of generation. R is an object-oriented programming language. This means that everything what is done with R can be saved as an object. Every object has a class. It describes what the object contains and what each function does. This is a review based on COVID 19 patients data on accessible big-data systems that include a set of tools and technique

REFERENCES

[1] C.L. Philip Chen, Chun-Yang Zhang, “Data intensive applications, challenges, techniques and technologies: A survey on Big Data” Information Science 0020-0255 (2014), PP 341-347, elsevier

[2] Han hu1At. Al. (Fellow, IEEE),” Toward Scalable Systems for Big Data Analytics: A Technology Tutorial”, IEEE 2169-3536(2014),PP 652-687

[3] Shweta Pandey, Dr.VrindaTokekar,” Prominence of MapReduce in BIG DATA Processing”, IEEE (Fourth International Conference on Communication Systems and Network Technologies)978-1-4799-3070-8/14, PP 555-560

[4] Katarina Grolinger At. Al.“Challenges for MapReduce in Big Data”, IEEE (10th World Congress on Services)978-1-4799-5069-0/14,PP 182-189

[5] Zhen Jia1 At. Al.“Characterizing and Subsetting Big Data Workloads", IEEE 978-1-4799-6454-3/14, PP 191-201

[6] AvitaKatal, Mohammad Wazid, R H Goudar, “Big Data: Issues, Challenges, Tools and Good Practices”, IEEE 978-1-4799-0192-0/13,PP 404-409

[7] Du Zhang,” Inconsistencies in Big Data”, IEEE 978-1-4799-0783-0/13, PP 61-67

[8] ZibinZheng, Jieming Zhu, and Michael R. Lyu, “Service-generated Big Data and Big Data-as-a-Service: An Overview”, IEEE (International Congress on Big Data) 978-0-7695-5006-0/13, PP 403-410

[9] VigneshPrajapati, Big Data Analytics with R and HadoopPackt Publishing

[10] Lei Wang At. Al., “BigDataBench: aBigDataBenchmarkSuitefromInternetServices”,IEEE 978-1-4799-3097-5/14.

[11] AnirudhKadadi At. Al., “Challenges of Data Integration and Interoperability in Big Data”, IEEE (International Conference on Big Data)978-1-4799-5666-1/14, PP 38-40

[12] SAS, Five big data challenges and how to overcome them with visual analytics

[13] HajarMousanif At. Al., “From Big Data to Big Projects: a Step-by-step Roadmap”, IEEE (International Conference on Future Internet of Things and Cloud) 978-1-4799-4357-9/14, PP 373-378

[14] Tianbo Lu At. Al., “Next Big Thing in Big Data: The Security of the ICT Supply Chain”, IEEE (SocialCom/PASSAT/BigData/EconCom/BioMedCom) 978-0-7695-5137-1/13, PP 1066-1073

[15] Ganapathy Mani, NimaBarit, Duoduo Liao, Simon Berkovich, “Organization of Knowledge Extraction from Big Data Systems”, IEEE (4 Fifth International Conference on Computing for Geospatial Research and Application) 978-1-4799-4321-0/14, PP 63-69

[16] Joseph Rickert, “Big Data Analysis with Revolution R Enterprise”, 2011

[17] Carson Kai-Sang Leung, Richard Kyle MacKinnon, Fan Jiang, “Reducing the Search Space for Big Data Mining for Interesting Patterns from Uncertain Data”, IEEE 2014, PP 315-322

[18] Ajith Abraham1, Swagatam Das2, and Sandip Roy3, “Swarm Intelligence Algorithms for Data Clustering”, PP 280-313

[19] Swagatam Das, Ajith Abraham, Senior Member, IEEE, and Amit Konar, “Automatic Clustering Using an Improved Differential Evolution Algorithm”, IEEE 2008, PP 218-237

[20] Rodriguez-Morales AJ, Bonilla-Aldana DK, albin-Ramon GJ, Rabaan AA, Sah R, Paniz-Mondolfi A, Pagliano P, Esposito S. 2020. History is repeating itself: Probable zoonotic spillover as the cause of the 2019 novel Coronavirus Epidemic. Infez Med 28(1):3-5.

[21] Gralinski LE, Menachery VD. 2020. Return of the Coronavirus: 2019-nCoV. Viruses 12(2):E135. doi: 10.3390/v12020135.

[22] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus Investigating and Research Team. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 10.1056/NEJMoa2001017. doi: 10.1056/NEJMoa 2001017.

[23]Wei X, Li X, Cui J. 2020. Evolutionary perspectives on novel Coronaviruses identified in pneumonia cases in China. National Science Review. doi: 10.1093/nsr/nwaa009.

[24]Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. 2020. A novel Coronavirus emerging in China-key questions for impact assessment N Engl J Med 10.1056/NEJMp2000929. doi:10.1056/NEJMp 2000929

[25] Fan Y, Zhao K, Shi ZL, Zhou P. 2019. Bat Coronaviruses in China. Viruses 11(3):210. doi:10.3390/v11030210.

[26] Lu H. 2020. Drug treatment options for the 2019-new coronavirus (2019-nCoV). Biosci Trends 10.5582/bst.2020.01020.doi: 10.5582/bst. 2020.01020.