Chapter 81
Intelligent Big Data Analytics in Health

Ebru Aydindag Bayrak

https://orcid.org/0000-0002-2637-9245
Istanbul University – Cerrahpaşa, Turkey

Pinar Kirci
Istanbul University – Cerrahpaşa, Turkey

ABSTRACT

Intelligent big data analytics and machine learning systems have been introduced to explain for the early diagnosis of neurological disorders. A number of scholarly researches about intelligent big data analytics in healthcare and machine learning system used in the healthcare system have been mentioned. The authors have explained the definition of big data, big data samples, and big data analytics. But the main goal is helping researchers or specialists in providing opinion about diagnosing or predicting neurological disorders using intelligent big data analytics and machine learning. Therefore, they focused on the healthcare systems using these innovative ways in particular. The information of platform and tools about big data analytics in healthcare is investigated. Numerous academic studies based on the detection of neurological disorders using both machine learning methods and big data analytics have been reviewed.

INTRODUCTION

The concept of big data was first used by Michael Cox and David Ellsworth at Proceedings of the 8th Conference on Visualization held in 1997, entitled "Application Controlled Demand Paging for Out-of-core Visualization". In the same study, it was mentioned that the datasets were too big and the computer system filled up the memory, disks and even external disks, and this problem was called "Big Data Problem" (Aktan, 2018).

The term big data was used for using larger volumes of scientific data for visualization. Although there are a large number of definitions of big data in the literature, the most popular definition comes from IBM. Big data could be characterized by any or all of three "V" words as suggested by IBM. V means that volume, variety, and velocity (O'Leary, 2013).

2.5 quintillion bytes of data was created by people, that is to say ninety percent of data (%90) has just been created in the last two years. This data is generated social media posts, videos, cell phone GPS signals or sensors. Here it is, this data is called Big Data (IBM, n.d.a.).

According to Gartner Incorporation "Big data is high-volume, high-variety and/or high-velocity information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation" (Gartner IT Glossary, n.d.).

The concept of big data; can be defined as a problem that occurs when traditional database management systems are inadequate when the data is stored, analyzed and managed (Sağıroğlu, 2017).

Big data indicate to growing dataset that involves unstructured, structured and semi-structured data by contrast with traditional data. The term big data was defined using the three main characteristics (3V) by most scientists and experts (Oussous, Benjelloun, Lahcen and Belfkih, 2017).

- Volume: It means the size of data which is varying from different data unit (terabyte, petabyte). Digital devices and applications (smartphones, IoT, social networks, logs,…) are generate big volumes of digital data. According to the report of International Data Corporation (IDC) the volume of data will increase from 898 exabytes to 6.6 zettabytes between 2012 and 2020. In other words, data will grow more than % 25 per a year.
- Variety: Big data is a variety of different formats (logs, videos, sensors,…) and sources. So it means the diversity of datasets.
- Velocity: Data is generated in a fast way that is means speed of data change.

The three components of big data can be summarized as in Figure 1. In addition to the three 3V's, other dimensions of big data have also been mentioned. These include (Gandomi and Haider, 2015):

- Veracity: This concept was coined by IBM to represent the uncertainty in some sources of data. We can give example such as customer sentiments in social media that are uncertain and include personal opinion. Even so they are valuable for analyzing information.
- Variability: It refers to the variation in the data flow rates, was introduced by SAS (Statistical Analysis Software).
- Value: It is defined by Oracle to define attribute of big data. Clearly it can be explained creating a value to organizations using big data analysis in the decision-making.

When research is done in both academic and business literature Big Data has been identified four key themes to which refers: Information, Technologies, Methods and Impact (De Mauro, Greco & Grimaldi, 2015).

The evaluation of big data is explained that it is equivalent to the oil of 20th century and is the gold mine of the 21th century. It is valuable for organization, government and individual (Sun, 2017).

BACKGROUND

O'Leary (2013) have focused on some of the basic concern and uses of artificial intelligence for big data. About the integration of artificial intelligence and big data case studies were presented. As what is big data, the application of mapreduce and hadoop, the significance of structured data topics were

38 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/intelligent-big-data-analytics-in-health/291058

Related Content

Improvement in Task Scheduling Capabilities for SaaS Cloud Deployments Using Intelligent Schedulers
Supriya Sawwashere (2021). *International Journal of Big Data and Analytics in Healthcare (pp. 1-12).*
www.irma-international.org/article/improvement-in-task-scheduling-capabilities-for-saas-cloud-deployments-using-intelligent-schedulers/287104

A Real Time Topological Map Matching Methodology for GPS/GIS-Based Travel Behavior Studies
Carola A. Blazquezand Pablo A. Miranda (2014). *Mobile Technologies for Activity-Travel Data Collection and Analysis (pp. 152-170).*
www.irma-international.org/chapter/a-real-time-topological-map-matching-methodology-for-gpsgis-based-travel-behavior-studies/113209

Big Data and Enterprise Applications
Ahmet Doanand Emin Sertaç Ar (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications (pp. 842-868).*
www.irma-international.org/chapter/big-data-and-enterprise-applications/291014

Prediction of Heart Disease Using Random Forest and Rough Set Based Feature Selection
Indu Yekkalaand Sunanda Dixit (2018). *International Journal of Big Data and Analytics in Healthcare (pp. 1-12).*
www.irma-international.org/article/prediction-of-heart-disease-using-random-forest-and-rough-set-based-feature-selection/209737

A Markov-Chain-Based Model for Group Message Distribution in Connected Networks
Peter Bajorskiand Michael Kurdziel (2020). *International Journal of Data Analytics (pp. 13-29).*
www.irma-international.org/article/a-markov-chain-based-model-for-group-message-distribution-in-connected-networks/258918