Chapter 21
Intelligent Techniques for Analysis of Big Data About Healthcare and Medical Records

Pinar Kirci
Istanbul University, Turkey

ABSTRACT

To define huge datasets, the term of big data is used. The considered "4 V" datasets imply volume, variety, velocity and value for many areas especially in medical images, electronic medical records (EMR) and biometrics data. To process and manage such datasets at storage, analysis and visualization states are challenging processes. Recent improvements in communication and transmission technologies provide efficient solutions. Big data solutions should be multithreaded and data access approaches should be tailored to big amounts of semi-structured/unstructured data. Software programming frameworks with a distributed file system (DFS) that owns more units compared with the disk blocks in an operating system to multithread computing task are utilized to cope with these difficulties. Huge datasets in data storage and analysis of healthcare industry need new solutions because old fashioned and traditional analytic tools become useless.

INTRODUCTION

In the work of Mayer-Schönberger and Cukier (2013), people have collected vast amounts of data for centuries in libraries since ancient times, thus collecting and accumulating data is not new for human race. Today, they want to keep whole of the considered data. Companies want to gather the data of their suppliers, customers and staff. Also, the data of business transactions, of purchases and sales, of expenses and profits for many years. Nearly twenty years ago, this kind of data were kept in books and files. But today it is impossible to keep this kind of valuable data in files because it is not safe and there is not enough space. Today, electronic storage on databases is utilized to keep data. Thus, people save time, space, money and effort with keeping the past and present data of the companies with larger databases, faster processing computers and greater storage capacities. Larger databases are provided by

ever improving computer technology. Databases are a kind of two-dimensional table of data and thus they grow in two ways as in the number of rows and the number of columns. These rows and columns keep various types of data including names, expiration dates, addresses, salaries of customers, goods, purchase and staffs. In time, the data entries increase but there is no need to remove older entries because today storage is not a problem, thus gathering and keeping data is easier. Besides, with one barcode scanner, much of the vital data can be easily gathered about a product such as transaction's time and date, payment type and customer's data. Also, internet presents large amounts of data, easy access, gathering and storing only with a click.

According to Mayer-Schönberger and Cukier (2013), in addition to the ever growing databases, many databases come together and form data warehouses. Nowadays, even small companies need data warehouses instead of using a database because of the variety of data needed. By connecting databases, it is easier to reach large amounts of data such as several petabytes, thus relationships are emerged among many variables and vast amounts of data can be received from data warehouses so that this will rise to data mining. But it is obvious that in a few years data warehouses will be insufficient because of the ascending data. Therefore, Cloud Computing is offered together with online storage of data with vast networks of servers. The data that cannot be handled by traditional database technology because of its unstructured nature or large amounts is called as big data. It is composed of three V's which are volume, velocity and variability. Velocity is the speed of the processes which are data collection, storage and analysis. It is easy to handle numerical data but video clips, music and text recognition including unstructured data is more complex and difficult for processing. Thus, sometimes veracity *named* fourth V emerges because truth is an important factor. In a day, Google handles nearly 24 petabytes of data, 3 billion comments are left in Facebook and 400 million messages are left on Twitter. This means that at every minute nearly five exabytes of data circulates over the world.

Obaidat et al. (2011) discuss that the security, efficiency, patient based, seasonableness, impartialness and efficiency over different nations are basic goals that should be ensured for the development of healthcare. These goals were mentioned by The American Institute of Medicine. A mobile data infrastructure that includes telemedicine systems and information processing techniques for patients' needs will be able to ensure these aims. Lately, wearable and pervasive sensors, working with personal mobile devices addresses new e-health systems which are pervasive healthcare systems. The system combines patient based and hospital based systems with ensuring medical workflow inside the facility with collecting more data.

The authors of Kenny (2014) and Obaidat et al. (2011) discuss in their research that with the technological developments, keeping, processing and storing data will be less expensive because of less space requirement. Sensor usage will provide more data collection and transmission to wherever the data is required. Monitoring of engines, people, environment and machinery will also make life easier with collecting more data. Besides, with the help of collecting private and health data by remote health monitoring improves the life quality of patients in their own homes. Personal Health Systems with wearable sensors gather, process, store and send medical data ensuring the life standards of patients by monitoring and communicating with external devices. All of these advancements will be provided with smaller devices for big data storage and processing. In the last ten years, for keeping a definite amount of data, smaller volume of space was needed. With the recent developments in nanotechnology, storage units are produced in the atomic level. Furthermore, increasing the data processing speed is another vital topic. Today, parallel arrangement of computers is used to improve the data processing speed.

Related Content

A New Algorithm for Mode Detection in Travel Surveys: Mobile Technologies for Activity-Travel Data Collection and Analysis
Birgit Kohla, Regine Gerike, Reinhard Hössinger, Michael Meschik, Gerd Sammerand Wiebke Unbehaun (2014). *Mobile Technologies for Activity-Travel Data Collection and Analysis (pp. 134-151).*
www.irma-international.org/chapter/a-new-algorithm-for-mode-detection-in-travel-surveys/113208

Comprehensive Analysis of State-of-the-Art CAD Tools and Techniques for Chronic Kidney Disease (CKD)
Mynapati Lakshmi Prasudha, Rakesh Kasumollaand Deepak Sukheja (2021). *International Journal of Big Data and Analytics in Healthcare (pp. 1-12).*
www.irma-international.org/article/comprehensive-analysis-of-state-of-the-art-cad-tools-and-techniques-for-chronic-kidney-disease-ckd/287605

Fuzzy Bayesian Context-Aware System to Reduce Electricity Consumption
Kavita Pankaj Shirsatand Girish P. Bhole (2021). *International Journal of Data Analytics (pp. 86-98).*
www.irma-international.org/article/fuzzy-bayesian-context-aware-system-to-reduce-electricity-consumption/272110

Implementation of Big Data Analytics for Government Enterprise
Namhla Matiwaneand Tiko Iyamu (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications (pp. 1294-1306).*
www.irma-international.org/chapter/implementation-of-big-data-analytics-for-government-enterprise/291038

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