U-healthcare Big Data Analytics Process Control

Mechelle Grace Zaragoza[1], Haeng-Kon Kim[2] and Younky Chung[3*]

[1,2]*School of Information Technology, Catholic University of Daegu, Korea*
[3]*Department of Computer Engineering, Kyungil University, Republic of Korea*
[1]*mechellezaragoza@gmail.com, [2]hangkon@cu.ac.kr, [3]ykchung@kiu.ac.kr*

Abstract

Big data analytics made its way beyond expectation due to its huge potential applied in various areas such as science and technology. As to healthcare is concerned, data is becoming more complex and is difficult to process in the long run. How to manage data has been an important issue to resolve. New development was also successful in collecting and storing healthcare data anytime remotely, hence the ubiquitous healthcare systems. Big data analytics has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions. In this study, we will discuss the use and application of big data analytics for u-healthcare. We have presented architecture to address the mentioned challenges in this study by improving the current process of data storage through encryption and setting access rights.

Keywords*: big data, u-healthcare, big data analytics*

1. Introduction

Big Data generally refers to data that exceeds the typical storage, processing and calculation capabilities of traditional databases and data analysis techniques. Historically, the healthcare industry has generated large amounts of data, guided by record keeping, compliance and regulatory requirements, and patient care. By definition, important data on health care refer to electronic health data sets so large and complex that they are difficult (or impossible) to manage traditional software and or hardware nor can they be easily managed with traditional or common data management tools and methods. Important health care data includes data on patient care, such as physician notes, lab reports, X-ray reports, case histories, social media, sensor data, diet, list of physicians and nurses in a particular hospital, national data on health record, identification of the expiration date of medical and surgical instruments based on RFID data. Important data on health care is overwhelming not only because of its volume, but also because of the diversity of data types and the speed with which it should be managed. All data on health care and patient well-being are "great data" in the healthcare industry. [1] Health care systems have become ubiquitous health systems (u-healthcare), which collect important data using sensors anywhere. When the patient's behavioral traits, attitudes, habits, basic information such as age, ethnicity, height, weight, symptoms and the cumulative clinical history, these data become complex, which may be a vital attribute in predicting the common causes of the disease to prevent the development of such diseases. Big Data generally refers to data that, by digitizing, combining and effectively using large data, healthcare organizations ranging from single-doctor offices and multi-vendor groups to large-scale networks hospitals and responsible care organizations can attain noteworthy
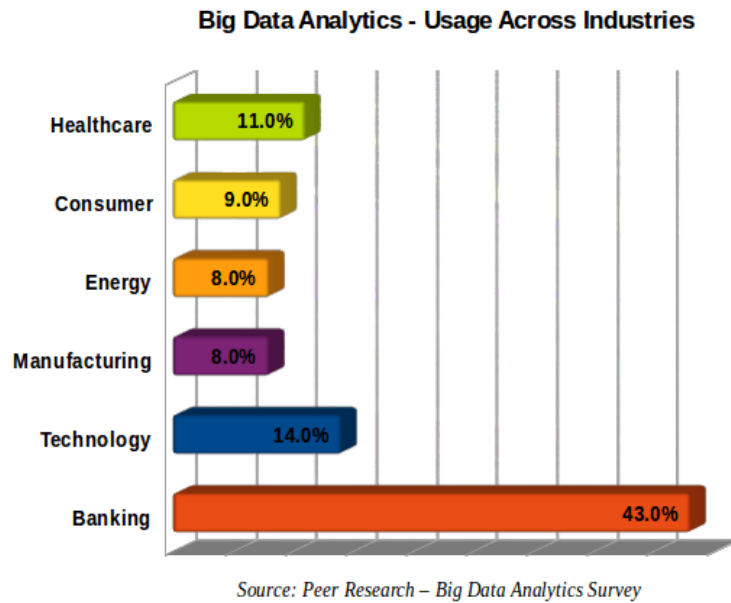
benefits.

Figure 1. Big Data Analytics Survey based in Industries

In Figure 1, it is a given depicts the job opportunities across various domains that there is a huge demand for Big Data Analytics owing to its awesome features. The tremendous growth is also due to the varied domain across which Analytics is being utilized.

Figure 2, Analysis is a key resource for many companies. There is no doubt about it. According to the "Analytics Advantage" survey overseen by Tom Davenport, ninety-six percent of respondents believe that the analysis will be more important to their organizations in the next three years.
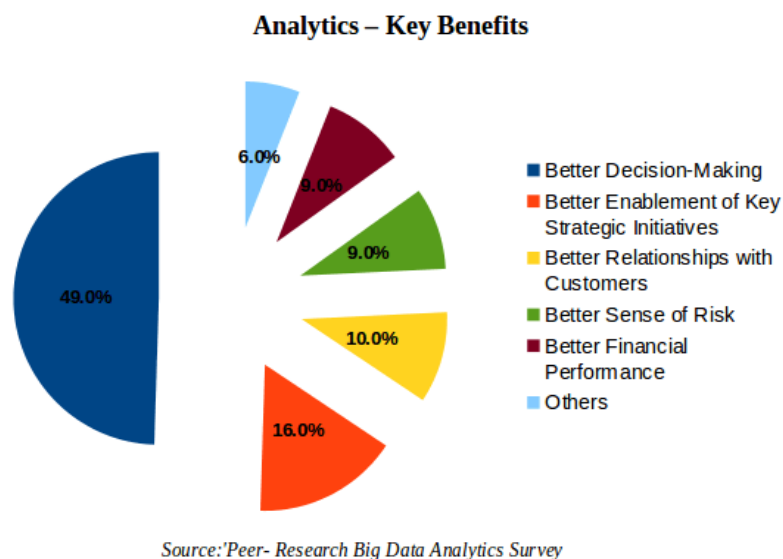

Figure 2. Big Data Analytics Key Benefits

This is because there is a large amount of data that is not being used and, at this stage, only rudimentary analyzes are being performed. About forty-nine percent of respondents firmly believe that analysis is a key factor in improving decision-making capabilities.

Another sixteen percent likes it for its main strategic initiatives.

Although there is a title fight for "Biggest Big Data Analytics Advantage", one thing is undeniable and stands out more: analytics play an important role in driving business strategy and effective commercial decision making. [2]

Potential benefits include early detection of disease when it can be treated more easily and efficiently; the management of individual and population health and the more rapid and effective detection of health care fraud is also possible.

Many problems can be addressed with a large data analysis. Some developments or outcomes can be predicted and/or estimated on the basis of large amounts of historical data, such as length of stay (LOS); patients electing elective surgery; patients who are unlikely to benefit from surgery; complications; patients at risk of medical complications; patients with risk or other hospitalized diseases; disease progression; patients at risk of advancement in pathological conditions; causal factors of disease / progression of disease.

2. Background of the Study

The volume of health data is expected to increase significantly in the coming years. In addition, health care reimbursement patterns change; significant use and payment for performance are emerging as critical factors in today's health environment. Although benefit is not and should not be a primary motivator, it is vitally important that health organizations acquire the tools, infrastructure, and techniques available to effectively exploit the great data or the risk of potentially losing millions of dollars in revenue and profits [3].

The enormous amount of medical and medical data currently untested can be utilized using existing analytical techniques that can lead to better understanding.

2.1. Big Data Analytics

The amount of digitally collected and stored data is huge and growing rapidly. As a result, the science of data management and analysis is also moving forward to enable organizations to turn this vast resource into information and knowledge that helps them achieve their goals. Computer scientists have invented the term great data to describe this evolution of technology. [4] Large data analysis can be defined as the process of examining large amounts of data from a variety of data sources and formats to provide real-time or real-time decision making information. Several analytical concepts such as data mining, natural language processing, artificial intelligence and predictive analysis can be applied to analyze, contextualize and visualize data. The analysis is the systematic use of data with business ideas to guide evidence-based decision making for planning, management, and measurement and learning; has been developed through applied analytical disciplines such as statistics, contextual, quantitative, predictive, and cognitive and other emerging models. In addition, the analysis uses descriptive and predictive or prescriptive models to obtain meaningful information and finally gain valuable knowledge of the data. Therefore, the analysis should not be considered as an individual analysis or as a step of analysis. Analysis in health care also exists in the same context.

There are five main "dimensions" for Big Data, usually called Five V's: Volume, Speed, Variety, Veracity, and Value. By understanding Big Data's 5Vs, we can use their power in research and solve real problems. So let's see how these dimensions can be applied to health care.

Volume: Health data is in terabytes and petabytes. These systems include information such as personal information, radiological images, personal medical records, 3D images, genomic readings and biometric sensors. Health systems may now have the potential to manage and analyze these complex data.

Variety: A variety of data is structured, unstructured and semi-structured. Structured information, such as clinical data, is easy to manipulate, store and analyze by machine. Most healthcare data, such as office medical records, medical notes, paper prescriptions, X-ray pictures and films, are neither structured nor semi-structured.

Speed: Most healthcare data is traditionally static paper files, X-ray films, and hyphens. The speed of mounting data increases with data representing regular monitoring, such as multiple daily measurements of diabetic glucose (or more continuous control by insulin pumps), blood pressure readings and EKG. Meanwhile, in many medical situations, constant real-time data (blood pressure monitoring, anesthesia monitors, bedside monitors, *etc*.) can mean the difference between life and death.

Truth: Some practitioners and researchers have introduced a fourth characteristic, truthfulness or a "given assurance." In other words, large data analyzes, and results are error-free and credible. Of course, truthfulness is the goal, not (yet) reality. Data quality issues are a major concern in health care for two reasons: decisions about life and death depend on accurate information and the quality of health service data; it is very variable and too often incorrect. (The "inaccurate translations" of the bad lyrics of the recipes are perhaps the most infamous example).

Value: Variety Unlike the 4V mentioned above, this V is too special because it represents the desired results of large data processing. We are always interested in collecting and extracting so much real value from the important data we work with.
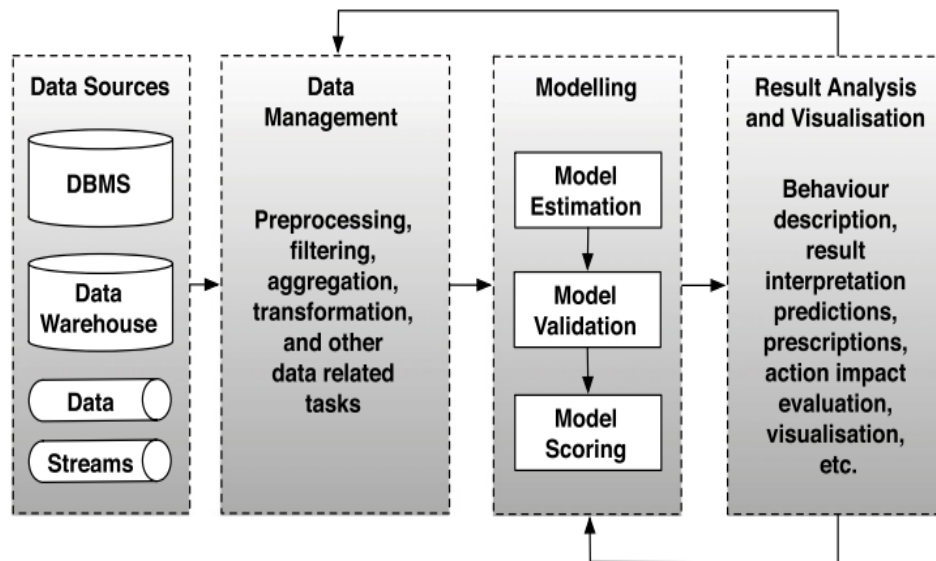

Figure 3. Overview of the Analytics Workflow for Big Data

In fact, we have to look for investments in data storage because the value of the data depends on the government's strategy and quality mechanism. For example, storing clinical data for new diseases in unreliable storage can save money today, but it can affect data tomorrow. [1]

Figure 3, represents the common phases of a traditional analysis workflow for Big Data. Data from various sources, including databases, flows, marts and data warehouses, are used to build Models. Large volume and different types of data may require pre-processing tasks for data integration, cleaning and the prepared data are used to form a

model and estimate their parameters. Once the model is estimated, it must be validated before consumption. Normally, this phase requires the use original input data and specific methods to validate the created model. Finally, the model is consumed and applied to the data that happen. This phase, called notation model, is used to generate predictions, prescriptions and recommendations. Results are interpreted and evaluated, used to generate new models or calibrate existing or integrated with pre-processed data. [5]

2.2. U-Healthcare

With the rapid growth of communication technologies, the collection and remote storage of health care data is easier and more affordable. This translates into a convenient and efficient service that allows users to work in locations and produce better results.

Ubiquitous healthcare applications may include disease diagnosis devices, surveillance systems, and even healthcare information systems. In omnipresent care, wireless data communication technologies such as WSN and Bluetooth® are adapted to transmit data when mobile devices are used as a surveillance terminal. However, several limitations and problems are reported with the adoption of Bluetooth® in ubiquitous communications in the field of health, such as energy needs. WSN being able to overcome problems seems to be more favorable for remote wireless health communications. In WSN, each sensor has a wireless communication capability and a certain level of intelligence for the processing of signals and data networks. In addition to direct transmission, a spatially distributed sensor has the ability to route data through a multi-hop routing protocol through a few sensor nodes to a gateway sensor node called node bases. This provides better connectivity than Bluetooth® technology, which only supports transmission to a certain maximum range. [6]

2.3. Challenges for Big data Analytics and u-healthcare

The large data analysis platform in healthcare must support the key functions required for data processing. Platform evaluation criteria may include availability, continuity, ease of use, scalability, multi-level granularity handling, privacy and security protection, and quality assurance.

- Include knowledge of complex and heterogeneous patient sources. Taking advantage of patient /data correlations in longitudinal registers.

- Understand unstructured clinical notes in the right context.

- Effective processing of large volumes of medical imaging data and extraction of potentially useful information and biomarkers.

- Genomic data analysis is a computationally intensive task and combining with standard clinical data adds additional layers of complexity.

- Capturing patient behavior data through the use of multiple sensors; their various interactions and social communications. [7]

- Confidentiality and sensitivity of information

The data source of the organizations (hospitals, pharmacies, companies, medical centers) is in different formats. These organizations have data in different systems and parameters. To use this huge amount of data, these organizations must have a common data warehouse in order to obtain consistent information and be able to handle it. However, these systems require enormous costs.

Data quality is a serious constraint. The data collected are, in some cases, unstructured, inadequate and non-standardized. As a result, the industry must make additional efforts to transform information into usable and meaningful data.

A large investment is required for companies to acquire personnel (scientific data), resources and also to purchase data analysis technologies. In addition, companies must convince medical organizations to use large data analysis. The use of data mining and analysis of large data requires a high level of knowledge and experience. It is an expensive business for companies to hire these people. Due to significant limitations in the quality of the data collected, variations and errors in the results are not excluded. The idea is to make people's health records and medical history easily accessible to them at any time, through the office or any mobile device. The existence of such an electronic health record (EHR) could save lives in many critical situations. In addition, an EHR offers the advantage of not having to store large copies of the whole family. Paper-based records are cheap, but have limitations such as difficult access, time to date, security, inability to share and maintain over a long life. The practice of electronic health care and medical records is evolving in developed countries such as the United States, the United Kingdom, *etc*. The country, like Australia, is dedicated to developing a permanent electronic health record for all its citizens. But the question of the viability of this idea arises in a country like India. This may be possible if all sectors of society, especially the poor, are supported by government [8].
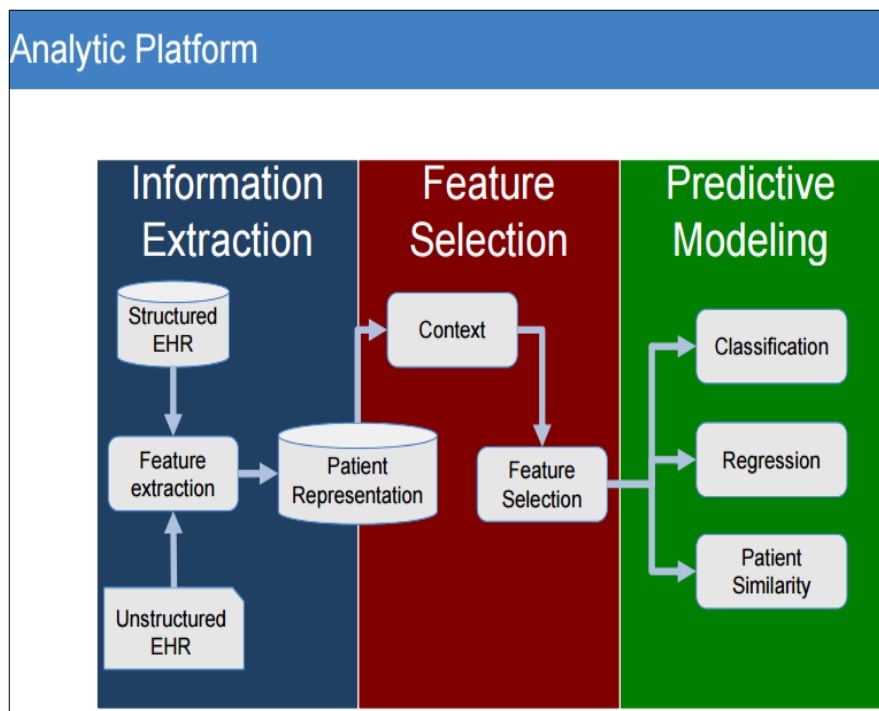

Figure 4. Big Data Analytic Platform [9]

The list below is not exhaustive, and as more research is done in this area, the more difficult questions will happen in Big Data?

Variety of data: how to manage an increasing volume of the data? Especially when the data is not structured, how to quickly extract meaningful content? How add and correlate broadcast data from multiple sources?

Data storage: how to recognize and store effectively information extracted from unstructured data? How store large volumes of information in a way that can be recovered in a timely manner? Are current file systems optimized for the volume and variety required by analytical applications? If not, what new capabilities are needed? How store information so it can be easily migrated / cover between data centers / cloud providers?

Data integration: new protocols and interfaces for integration data that can handle data of different nature (structured, unstructured, semi-structured) and sources.

Data Processing and Resource Management: New Programming models optimized for transmission and / or multidimensional The data; new optimized back-end engines file systems; engines capable of combining applications of various programming models (for example, MapReduce, workflows, and task bag) into a single solution / abstraction. How to optimize the use of resources and energy consumption when running the analysis application? [5]

3. Designed U-healthcare Data Analytics Process

Data on health care are becoming more complex. Great data analysis has the potential to transform the way healthcare providers use sophisticated technologies to better understand their clinical and other data repositories and make informed decisions. In the future we will see the rapid and widespread implementation and use of large-scale data analysis in the organization of health and the health industry. To this end, it is necessary to address the various challenges outlined above. As data analysis becomes more common, issues such as privacy, security, norms and governance, and continuous improvement of tools and technologies attention. Major data analysis and applications in health care are at an early stage of development, but rapid advances in platforms and tools can accelerate their maturation process. The application of a large data analysis in the Electronic Health Register (EHR) will be very beneficial for all.

Previous studies have been successful in using ubiquitous systems in the storage and processing of health care data. This study will encompass large data analysis, data cleansing, including inferred validations and data encryption. As presented in Figure 5.

Extraction of Information

This step encompasses not only the implementation of structure and unstructured data, but also transforms unstructured data into structured data. Once the data has been standardized, an additional step of removing duplicates can be included.

Feature selection

Data mining techniques are applied in this step to filter, sort, or group correctly. The data mining techniques below can be used to derive the desired data presentation.

- Exhaustive
- Best first
- Simulated annealing
- Genetic algorithm
- Greedy forward selection
- Greedy backward elimination
- Particle swarm optimization
- Targeted projection pursuit
- Scatter Search
- Variable Neighborhood Search

Predictive modeling

This data mining step is based on the probability of predicting the results using equations and formulas applicable in the statistical analysis.

Given the previous large data analysis process, other steps can be added below to address ethical issues related to data privacy.

Encryption and storage

- Storing data in a ubiquitous system, especially in cloud storage, can make data sensitive to unauthorized access, such as hacking. This additional step can encrypt the data to prevent unauthorized personnel from accessing or using the data. In order to encrypt the data, there is existing data encryption software that can be used, as required by the system. Some algorithm is used to encrypt the data, under some commonly used.
- DES/3DES or TripleDES
- Blowfish
- AES (Advanced Encryption Standard or Rijndael)
- Twofish
- IDEA
- MD5
- SHA 1
- HMAC
- RSA Security

Access Rights Management

In protecting Confidentiality, Integrity and Availability (CIA) access management can be set in place that will grant authorized users the right to use or access data while preventing access to non-authorized users. Layers of authorization can also be applied, such as multi-factor identification, making the data more secured.

Information Decryption

Once data is accessed by an authorized user, the encrypted data needed will be transformed back for the users disposition.
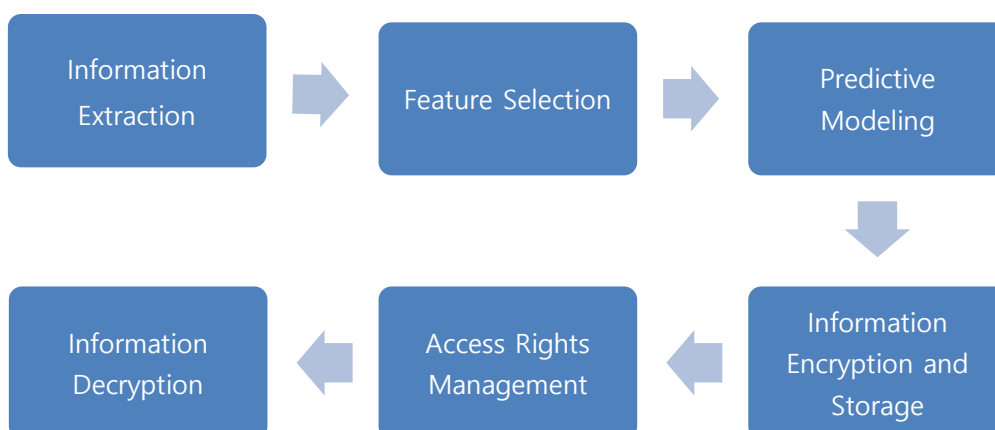

Figure 5. Proposed U-healthcare Data Analytics Process

4. Conclusion and Future Works

Data on health care are becoming more complex. Great data analysis has the potential to transform the way healthcare providers use sophisticated technologies to better understand their clinical and other data repositories and make informed decisions. Application of analytics in processing data for u-healthcare to come up with solutions or practices may not only improve the current pace of technology, but also improve the medical system not only of one area, population or country, but of the world. Certain standards or guidelines can also be set by the authorized agencies or personnel to address the ethical issues of accessing and using sensitive patient data. Future studies may include races, culture and ethnicities to determine and predict the underlying factors why diseases can affect certain population only.

Acknowledgements

References

[1]   V. Ganjir, B. K. Sarkar and R. R. Kumar, "Big data analytics for healthcare", International Journal of Research in Engineering, Technology and Science, Volume VI, Special Issue, **(2016)**.

[2]   D. Kaushik, "10 Reasons Why Big Data Analytics is the best career move", Edureka.co, **(2015)**.

[3]   W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential", Health Information Science and Systems (2014) https://www.biomedcentral.com/track/pdf/10.1186/2047-2501-2-3?site=hissjournal.biomedcentral.com.

[4]   A.S. Detsky and T. B. Murdoch, "The Inevitable Application of Big Data to Health Care", **(2013)** .http://jamanetwork.com/journals/jama/article-abstract/1674245.

[5]   D, M. Assunção, N. Rodrigo Calheiros, S. Bianchi, M.A. S. Netto and R. Buyya, "Big Data computing and clouds: Trends and future directions", Journal of Parallel and Distributed Computing, vol. 79, **(2015)**, pp. 3-15.

[6]   P. C. Hii and W. Y. Chung, "A Comprehensive Ubiquitous Healthcare Solution on an Android™ Mobile Device", **(2011)**.

[7]   I. Brown and A. A. Adams, "The ethical challenges of ubiquitous healthcare", http://fiz1.fh-potsdam.de/volltext/ijie/08242.pdf. **(2007)**.

[8]   J. Sun and C. K. Reddy, "Big Data Analytics for Healthcare", SIAM International Conference on Data Mining, Austin, TX, 2013. http://www.ijrets.com/wp-content/uploads/2016/12/16031408318.pdf.

[9]   R. Kavitha, E. Kannan and S. Kotteswaran, "Implementation of Cloud based Electronic Health Record (EHR) for Indian Healthcare Needs", **(2016).**

173