Big Data Framework for u-Healthcare System

Tae-Woong Kim [1], Jai-Hyun Seu [2]

[1.] Department of Computer Education, Silla University, Sasang-Gu, Busan, Korea
[2.] School of Computer Engineering, Inje University, Obang-dong 607, Gimhae, Gyeong-Nam, Korea
jaiseu@inje.ac.kr

Abstract: Recent u-Healthcare system can measure more complicated bio signal according to both built in sensor technology and personal carrying equipments' appearance. The system has been developed to analyze bio signal at anywhere and anytime. For example, there is a system to analyze abnormality and activity of body by measuring it for 24 hours a day. The system needs long time measured and accumulated bio signal, will have bigger capacity data and atypical data structure. In this paper, we define this bio signal as big data and propose u-Healthcare system framework which services the result of analyzed big data in real time. The proposed framework includes both bio signal transmission and analysis and real time service method. It uses open standard platform for all service process from bio signal transmission to final result of bio signal analysis to make machine compatibility and data interoperability. We verify framework of promising by servicing the result of analysis in real time which is collected object movement information by acceleration sensor.
[Tae Woong Kim, Jai Hyun Seu. **Big Data Framework for u-Healthcare System**. *Life Sci J* 2014;11(7s):112-116]. (ISSN:1097-8135). http://www.lifesciencesite.com. 20

Keywords: u-Healthcare; big data;

1. Introduction

U-Healthcare System is IT-based healthcare information system to prevent and help the early detection of the disease based on a number of vital signs, health status data from a person's body [1]. This will result they can forecast and manage their own health at anywhere anytime [2]. In the u-Healthcare System, the user's vital signs collection must be preceded. Then bio-signal analysis and vital signs processing, real-time response, biometric data storing technology should be applied on u-Healthcare System. Vital signs depending on the type of data collected are classified in two data types. The first data type is distinguishable simple data type like body temperature, blood pressure, blood sugar. Next is a continuous linear structure data type like electrocardiogram, respiration. These two types of data have different features. Blood pressure, body temperature, blood glucose was measured in any one moment, such as biometric data is suitable for storing in the database. Electrocardiogram, respiration, such as the linear structure of the biometric data is not suitable for storing the database structure [3]. Data types and volume differ from each other. Linear structure of the biometric data classified from the General data is needed to be handled as big data. There are a variety of analysis techniques for big data such as a reputation analysis, social network analysis, and etc but common used technique is Text Mining for big data analysis and statistics [4, 5]. But the vital sign data which has linear structure could not be analyzed directly. Therefore, we extract the values and characteristics from the vital sign data and convert it to the valid data format, and make it possible analysis data. Therefore, the big data platform for bio-signal analysis should have a different form of framework than a common big data platform for a common analysis and statistics. In General, there is Hadoop to implement the platform for big data [6]. Hadoop is not enough for a bio-signal analysis system or analysis of real time services system. Therefore, we define bio-data as big data and analysis these data for service. Finally we propose u-Healthcare System framework based on big data platform in this paper.

2. Big Data and Bio-signal

2.1 Big Data

Various kinds of large-scale data are being generated, distributed and stored rapidly in the midst of everyday life. Because of the existing social network service, which is represented by the growth of social media and the widely spread of mobile devices which is called smart phones. In addition, take advantage of usage extension for RFID and sensor devices used to detect information and the proliferation of cloud computing technologies to gather information is enabling the digital recording for personal information of the individual consumers and consumption behavior. Business Company and society need large-scale data processing, analysis and utilization technologies for forecast analysis based on the data collected [7]. Big data can be defined as a scale of dataset that goes beyond existing database management tool capabilities of data collection, storage, management, and analysis capabilities [4].

Big data shall have three properties 3V[3,4] such as Volume which is the huge amount of beyond existing data units, Velocity with the creation of the data flow is a very fast-pace and Variety of non-traditional structured data [Variety] mode, Photos, videos, and etc.

In addition, big data is classified as shown in table 1 depending on the severity of the formalized [6, 7, 8].

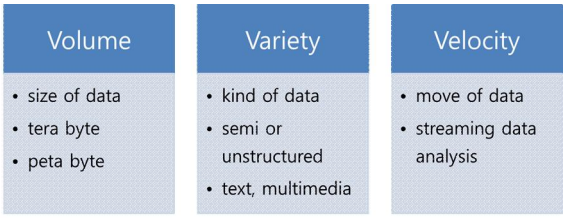

Figure 1. Three properties of BigData

Table 1. Defined Data Types in BigData

Definition	Explanation
Structured	Data Stored in fixed field. ex) Relational Database or spreadsheet
Semi-Structured	Data Stored not in fixed field but the data includes metadata, schema or etc. ex) XML or HTML text
Unstructured	Data not Stored in fixed field. ex) text document which is be able to be analyzed or image/video/voice data

Big data analysis techniques and methods are applicable to data mining, machine learning, natural language processing and pattern recognition which are used in existing statistics and computer science. In particular, text mining, opinion mining, cluster analysis and social network analysis has received attention due to the increase in atypical data such as social media (Table 2).

Table2. BigData Analysis technology

BigData Analysis technique	Explanation
Text Mining	The purpose of this technology is to process and extract useful information from Unstructured/Semi-formal type text data based on natural language processing technique.
Opinion Mining	A technology for determining preference for positive, negative, and neutral rating of structured/unstructured social media text.
Social Network Analysis	A technology to measure user's prestige and influence based on social networks the connectivity structure and connection strength, etc.
Cluster Analysis	A technology used to find out variation characteristics group by joining objects from objects with similar characteristics

2.2 Bio-Signal

Vital signs measurement technology is a critical measurement technology because vital signs are measured in very small sizes. There is a oxygen saturation sensor, acceleration sensor, blood pressure sensor, electrocardiogram sensor, body temperature sensor, respiration sensor to measure vital signal [11]. There are two types of data depending on type of biological signal is collected. One type is the measure value for a moment like body temperature, blood pressure, blood sugar and the other is subsequently entering data with a linear structure such as electrocardiogram, respiration, and acceleration signal [Figure 2]. These two types of data are very different on ways to be stored and analyzed. The simple distinguishable values are well suited for database storage but linear structured biometric data does not fit in database and the data-type and volume.

2.3 Bio-signal as Big Data

A series of linear structure vital signs data itself could not be considered to determine the status of the body. There is a need to go through the appropriate data processing. In order to represent the State of the body with proper bio-signal algorithm, we need to extract the value through the feature. Therefore, the feature value extraction algorithm should be included when the linear structure bio-signal data is analyzed. In addition, the extracted biometric signal feature values can be stored by using a common database.

However, these methods do not support data interoperability. Because you need to know the database schema in order to access the value of the extracted features, and every database type and each of the vendors have different features. Therefore, the

proposed framework uses the standardized format based on HL7 [Health Level Seven] [12] to store characterized value of biological signal [Figure 3]. This means that extended HL7 aECG [annotated ECG] [13] standard can be expressed in a variety of vital signs.

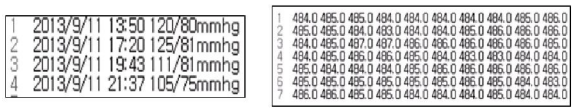

Figure 2. A simple feature value bio data

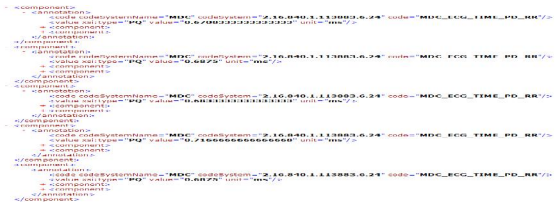

Figure 3. A Part of HL7 aECG structure

Vital signs are very meaningful if they are measured for a long time and have accumulated a period of time. Thus, the collected data is irregular, and its capacity is very large. If you look at the amount of data in semantic perspective it is desirable to classify it as big data. Therefore, it is quite natural defining these vital signs as big data and applying it on the u-Healthcare system. However, currently big data platform Hadoop is not easy to analyze vital signs. Because the problem how you collect data, whether the feature value extraction algorithm for vital sign mining is included or not, results of the analysis in real-time delivering services to consumers is due to the presence or absence of the system. In order to address these issues in this paper, we propose bio-signal analysis for big data framework that extends the Hadoop platform.

3. Big Data Platform for Bio-signal analysis

In recent years, the big data technologies are being developed to implement based on Hadoop from free Apache software. Hadoop stores a large amount of data on distributed storage and does parallel processing on multiple servers in a cluster. The Hadoop distributed file system, which operates in two separate nodes name nodes and data nodes, offers fast storage and stable. Map/Reduce as a subproject of Hadoop is the programming model for processing data, and simple but involve constitutional concurrency mechanism as shown in Figure 4. In this paper, we accommodate the storage models and data-processing model which Hadoop provides. We offer a bio-signal analysis extended framework as shown in Figure 5. In this framework, we add value storage model, the algorithm that extracts the value from Vital signs and SOA for real time services. In the framework, we use open standard SOAP message of the W3C for analysis result and vital sign transmission. Then the value data extracted by the signal analysis algorithm for interoperability will be stored based on HL7 standard Meta model.

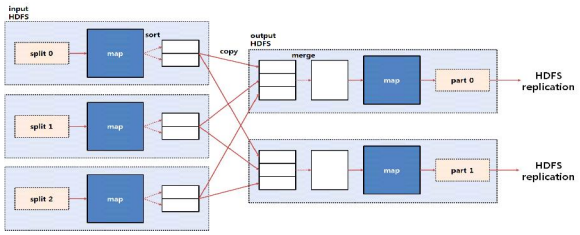

Figure 4. Map/Reduce Data processing mechanisms

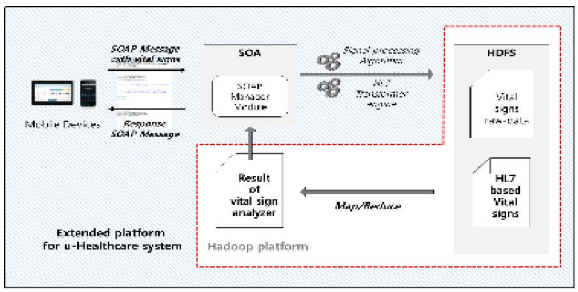

Figure 5. Extended BigData Framework for bio-signal

4. Implementation

This section uses the framework proposed in section 3 to implement. Vital signs can be extracted from a person's movement signals by acceleration sensors. Because the electrocardiogram, respiration requires a separate medical sensors but acceleration sensor is built into the commonly used Smartphone. Therefore it's easier to get the signal.

4.1 Bio-signal transmission

This process is sending the acceleration signal stored on Smartphone to the server. We use SOAP messages, which are the W3C standard RPC [Remote Procedure Call] models for heterogeneous device compatibility, because transfer devices may vary among users. HTTP is responsible for sending SOAP messages and we use SOAP Messages with Attachment to send the acceleration signal to the server. Figure 6 shows a part of SOAP message to send the acceleration signal to the server that includes the header and the actual acceleration of the signal raw data.

4.2 Feature value extraction algorithm

Acceleration signals have a value of 3 axes x, y, and z. The body's movements can be detected by the value of 3 axes, such as motion detection algorithm is much open to the public. In this paper, we explain the way how the extracted feature value is

stored in the proposed framework and will take advantage of the technology in the analysis. We can calculate steps, stride, calorie consumption, and etc by using the information extracted by motion detection algorithm in combination with an individual's height, weight. This algorithm is powered up and extracts the feature value and passes the resulting value to the Meta model-based vital signs feature value conversion module HL7 as soon as the acceleration signal transfer is completed from personal device.

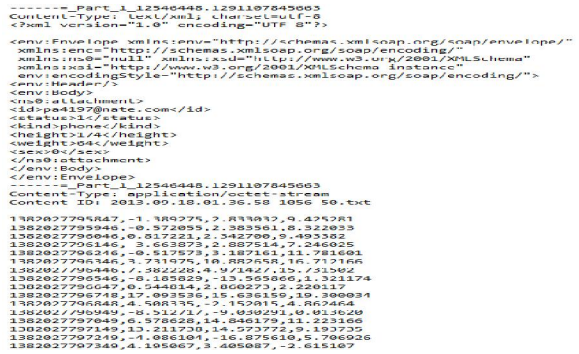

Figure 6. SOAP message for acceleration signal transmission

4.3 Bio data storage model

The feature value, extracted by acceleration signal analysis algorithms, is expressed in the HL7 standard aECG extension along with semantic information rather than storing in plain text or database for interoperability [Figure 7]. And the raw data is stored separately in HDFS. Originally aECG model defined by HL7 for expressing electrocardiogram is designed to contain raw data. But it is awkward to put a huge amount of raw data into an XML document for performance and big data structure. Raw data is inappropriate to approach semantically in big data. Therefore, this paper proposes a framework that stores a feature value and the raw data separately.

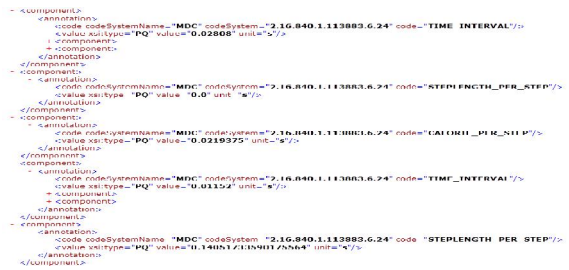

Figure 7. Acceleration feature value expression based on HL7

4.4 Analysis model

Generally in Hadoop, text mining techniques using a plain text file are primarily used for data analysis. But our data to analyze is an XML document based on the HL7. This means that a separate XML parsing should be used. By default, Hadoop provides programmable subproject called Map/Reduce for the text-mining. This makes valid data used for analysis as the key-value pair. We propose a framework that also represents key-value pair for the feature value. Therefore, we are using a Map/Reduce provided by Hadoop platform. It extracts steps per time, distance, and calorie consumption in the form of a key-value pair in map, and calculates the results values for final service items by using them in the list form of key-value pair in reduce. Figure 8 shows that these processing steps.

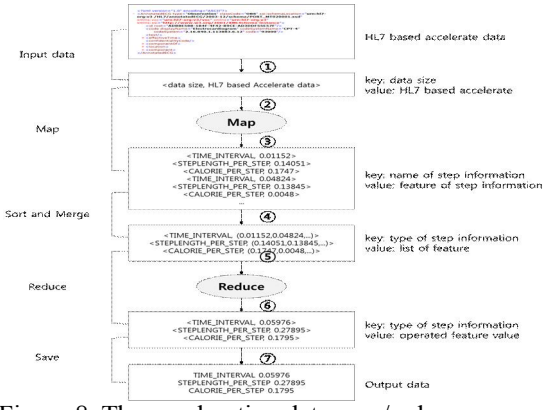

Figure 8. The acceleration data map/reduce processing

4.5 Service model

The result calculated by map/reduce is converted to SOAP message by SOA and send it to the clients [device]. SOA has the advantage that you can parse data regardless of the device type and OS as an open standard platform by using a W3C standard SOAP message for calling a heterogeneous object. Figure 9 shows the final SOAP message received by device that has collected and sent the acceleration signal to the server. It will be serviced the analysis result about the acceleration signal if you parse. The results obtained by analyzing the acceleration signals are calorie consumption, moving distance, number of steps, exercise hours and etc.

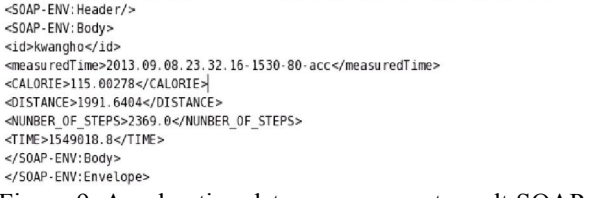

Figure 9. Acceleration data measurement result SOAP Message

5. Results

Biometric data is an important indicator for determining health status in U-Healthcare systems. Therefore, the classification, storage and analysis of the biometric data are important factors to judge performance and reliability of the u-Healthcare system. But the current biometric data storage and analysis technology do not consider features of biometric data, sensor dependency or signal dependent on the measuring sensor or development vendors have different features. This means the absence of u-Healthcare system standardization, the lack of data interoperability and systems integration problems. And unreliable biometric data store is inherent in such a degrading system performance and reliability of u-Healthcare. In this paper, we propose and implement bio-signal analysis system based on big data in order to solve these problems. We have a real-time bio-signal analysis for the expansion with features of existing big data solution, Hadoop platform. This means extension of the way to transmit vital signs to the server, the feature value extraction, data representation model and the real-time service model. And the proposed framework is implemented to collect acceleration signals on your Smartphone.

Corresponding Author:
PhD. Jai-Hyun Seu
School of Computer Engineering,
Inje University
Obang-dong 607, Gimhae, Gyeong-Nam, Korea
E-mail: jaiseu@inje.ac.kr

References

1. Pantelopoulos, and Nikolaos G, "A survey on wearable sensor-Based Systems for health Monitoring and Prognosis," IEEE Tran Sys Man and Cybernetics, vol.40(1), pp.1-12, 2010.
2. Victor J. Schoenbach, "Appraising health risk appraisal," Am J Public Health, 77, pp.409, 1987.
3. T.W.Kim and H.C.Kim, "A healthcare system as a service in the context of vital signs: Proposing a framework for realizing a model," Computers and Mathematics with Applications, Vol.64(5), pp.1324-1332, 2012.
4. James Manyika, Michael Chui, Brad Brown et al., "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, pp.15-17, 2011.
5. Magoulas, Roger, and Ben Lorica, "Big data: Technologies and techniques for large scale data," Release 2.0, Nov. 2009.
6. (2013) ApacheTM Hadoop® website. [Online]. Available: http://hadoop.apache.org/
7. John Gantz and David Reinsel, "Extracting Value from Chaos," IDC IVIEW, pp.6, 2011.
8. Health, T., and Bizer, C., "Linked data: Evolving the web into a global data space," San Rafael CA:Morgan & Claypool, 2011.
9. Hilbert, M., and Lopez, P., "The world's technological capacity to store, communicate, and compute information," Science, 332(6025), pp.60-65, 2011.
10. Lavigne, V., and Goulin, D., "Applicability of visual analytics to defense and security operations," Proceedings of the 16th International Command and Control Research and Technology Symposium, 2011.
11. Jun-Su Kim, Kayoung Lee, Cheol Seung Yoo et al., "Wearable Physiologic Monitoring System in Health Promotion," Korean Journal of Health Promotion, 11(1), pp.1-8, 2011.
12. (2013) Health Level Seven International website. [Online]. Avaliable: http://www.hl7.org/about/hl7about.htm
13. Barry D. Brown, Fabio Badilini. (2013) HL7 aECG Implementation Guide homepage. [Online]. Available: http://www.hl7.org.

4/28/2014