Curating Big Data for Health and Wellness in Cloudcentric IoT

저자 (Authors)	무하마드 비랄 아민, 무하마드 사딕, 막불 알리, 방재훈

출처 (Source)	한국통신학회지(정보와통신) 35(2), 2018.1, 42-57 (16 pages) The Journal of The Korean Institute of Communication Sciences 35(2), 2018.1, 42-57 (16 pages)
발행처 (Publisher)	한국통신학회 Korea Institute Of Communication Sciences
URL	http://www.dbpia.co.kr/Article/NODE07337155
APA Style	무하마드 비랄 아민, 무하마드 사딕, 막불 알리, 방재훈 (2018). Curating Big Data for Health and Wellness in Cloudcentric IoT. 한국통신학회지(정보와통신), 35(2), 42-57.

Curating Big Data for Health and Wellness in Cloud-centric IoT

무하마드 비랄 아민, 무하마드 사딕, 막불 알리, 방재훈
경희대학교

Abstract

In the era of IoT-based health and wellness, the dependence of decision making relies on the abundance of heterogeneous data generated by many small-scale sensory devices. The effective utilization of these IoT devices require collaborative execution with remote storage and computing power. Data acquisition from these devices is large in volume, heterogeneous and considered in the category of big data. This data is accumulated at scalable resources like cloud where intelligent processes constitute the foundation of smart environments. In this paper, we survey the state-of-the-art techniques, tools, and methodologies, implemented to support the curation of big data from IoT devices over cloud platforms.

I. Introduction

In recent years, the models of the conventional healthcare systems have changed. These models are now focusing on preventive yet personalized health. As a result, focus is now on when, where, and how; care and support should be provided to the consumer [1][2]. To achieve the goal of personalized-preventive healthcare Internet of Things (IoT) is a major contributor. As per concept of IoT, a cluster of homogeneous and heterogeneous devices works incoherently to accomplish a goal[3]. This accomplishment heavily depends upon the data accumulated from the sensory devices. Intelligent processes are applied on the collected data for context awareness and decision making[4]. For continuous sensing of data from sensory devices, cloud delivers the centralized hub of data gathering[5]. Consequently, Cloud-centric IoT [6] is now the research focus of distributed systems development and research. Furthermore, cloud being the centralized hub of data, facilitates processes to map user context over a timeline called lifelog[7]. Moreover, to support the volume of data and scalability of the platform, cloud provides big data storage and performance-based compute instances for provisioning.

With the focus on IoT and usability of cloud in the above-mentioned regards, an opportunity has emerged for researchers to allow people to take care of their health and wellness by providing them with timely, ubiquitous, and personalized support[8]. Therefore, the arrival of fitness wearables with smartphone applications and systems supporting IoT-based health and wellness has taken the market by storm[9]. Moreover, researchers are also working on health and wellness systems that can accumulate and monitor data from IoT and alert for physical conditions [10] or detect chronic illnesses [11]. Accumulated data being asset, points to a single goal of generation of context-rich user lifelog which can provide a holistic view of user activity and behavior[7]. Such lifelogs are essential for the evolutionary wellness systems that support the self-quantifications of its users. Furthermore, a context-rich lifelog is also a low-cost mean to obtain valuable user data on which advance methodologies like AI, descriptive, and predictive analytics can be applied for effective interventions from healthcare professionals[12].

In this paper, we are reviewing the state-of-the-art tools, techniques, and methodologies regarding the curation of big data originated from IoT devices over

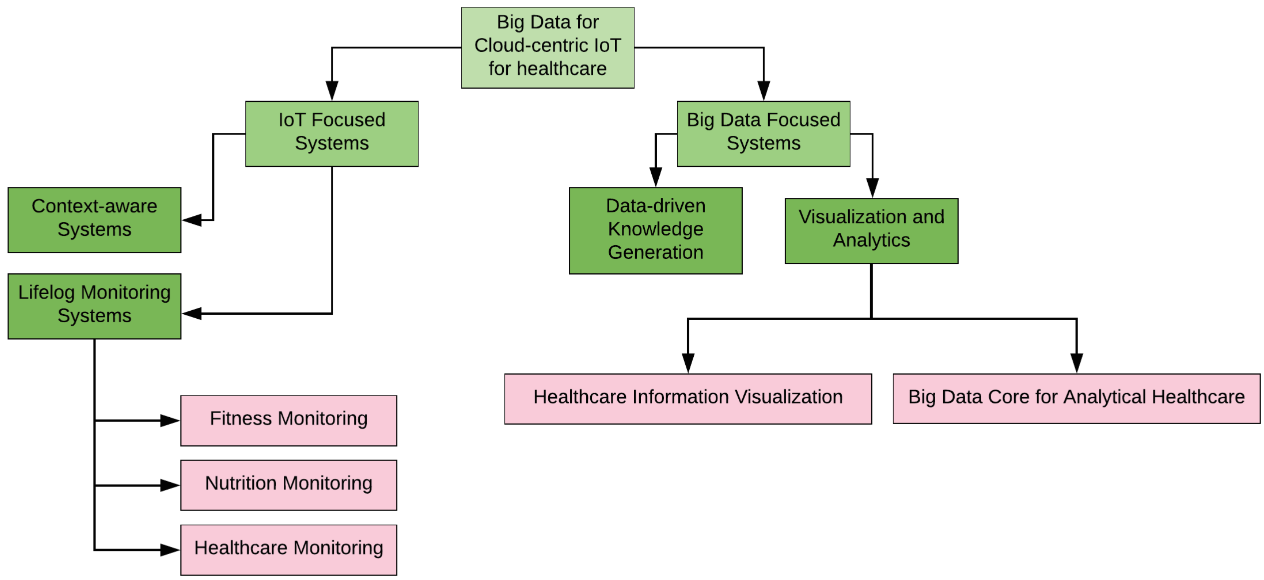


Fig. 1. Taxonomy

a cloud platform. Curation of data in this perspective covers the processing and persistence of the originated data. Furthermore, we are focusing on strategies implemented in the wellness domain more than clinical, as wellness domain has been the area of interest for many researchers and developers around the world. Moreover, the utilization of IoT in the domain of wellness is substantially more heterogeneous and hybrid than the Clinical domain.

As shown in the taxonomy 〈Fig. 1〉, we have divided cloud-centric IoT big data systems in two broad classifications, i.e., IoT focused and Big Data focused. IoT focused systems include implementations pertaining to stream of data from the point of origination till its persistence in a data storage. This storage can be of relational or non-volatile nature. Implementations in these classifications include context-aware systems determining low- and high-level user context from the IoT data stream. This classification also includes Lifelog Monitoring systems which process the stream of data looking for anomalies and situations. These systems respond as triggers in case of a situation detection. Foundation of these systems include comprehensive rule-based processes with parallel Just-In-Time (JIT) monitoring. These systems can also be classified as soft real-time processes.

On the other hand, Big data systems include implementations pertaining to processes applied over data storage. Storage in this perspective is most likely an unstructured big data source. Implementations in this classification includes visualization and analytics based systems where stored data is visualized based on queries generated by the user. These visualization systems incorporate techniques like predictive and descriptive analytics for reporting. Visualization and Analytics in Big data systems can be a soft real-time or offline implementation depending upon its use cases. This classification also includes Data-driven knowledge generation systems where stored big data is utilized in conjunction with Machine Learning techniques. These systems generally assist the expert for generating rules for the knowledge bases depending upon the historic data of a certain domain. This data is also used for the training of machine learning models. Data-driven knowledge base generation systems lie under the category of offline systems.

The general architecture of a Cloud-centric IoT system is illustrated in the 〈Fig. 2〉. The IoT is cluster of things where devices generate data as per their instructions. In a homogenous cluster of things, sensory devices

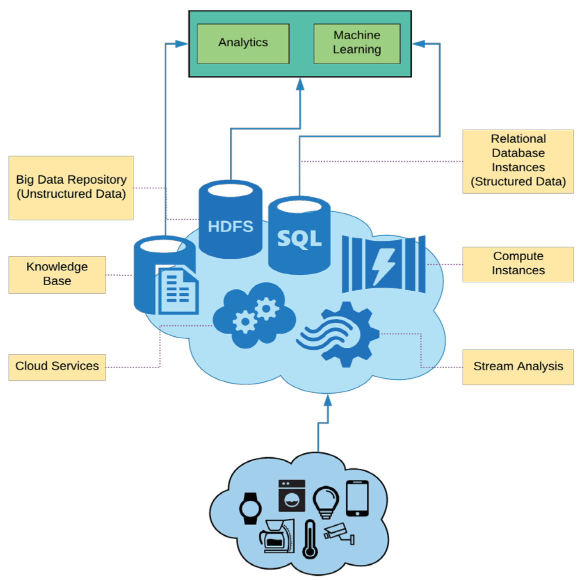


Fig. 2. Cloud-centric IoT general architecture

communicate among themselves based on machine-to-machine (M2M) communication protocols. In most of these implementations one device act as a head node responsible for communicating the collective data from the cluster of things to the cloud. However, in case of heterogeneous cluster of things, sensory devices communicate directly with the cloud. Most commonly used protocol of communication for these device is internet using http and https.

For the storage of structured data, RDBMS is the obvious choice. For rules, Knowledge base storage systems are provided. Availability of the storage depends upon the type of knowledge representation. For example, in case of large bio-medical ontologies, an implementation of a triple store or an object-oriented database may be provided. To store unstructured data, large storage platforms like HDFS for big data are provided. These data stores are categorized as non-volatile storage as data is not removed from persistence. Data access is provided using querying frameworks such as Map-reduce and Hive. Depending upon the computation needs, cloud provides compute instances. The strength of these instances depends upon the hardware support provided. For example, to support a compute-intensive problem, CPU with higher number of cores and larger memory is provided. These compute instances are provided with or without the support of an operating system. For performance needs, these compute instances can also provision fast storage such as Solid-State Drives (SSD) over conventional hard drives.

Utilization of cloud, depends upon the use cases and service scenarios. Data persisted in the cloud is commonly utilized in Post-persistence use cases such as descriptive and predictive analytics, or machine learning.

Although there are many generic cloud-centric implementations available; however, we are focusing on health and wellness domain for this survey. This paper is structured as follows: Section II surveys implementations under the umbrella IoT focused systems. Context-aware and lifelog monitoring systems are reviewed in this section. Section III surveys implementations under the umbrella of Big data focused systems. Visualization with analytics and data-driven knowledge generation implementations are reviewed in this section. In section IV, we provide details regarding our Cloud-centric IoT implementation called Data Curation Framework (DCF) for health and wellness systems[13]. Section V concludes this paper.

Ⅱ. IoT Focused Systems

1. Context-aware Systems

The widespread of sensors-handled mobile devices have accelerated the expansion of the IoT driven healthcare services in the smart environment. The introduction of IoT in the healthcare service industry provides innovative opportunities to improve not only healthcare but also guide for future research directions in pervasive sensing [14]. The ubiquitous identification, sensing, modeling and communication facilities for all IoT based objects like people, devices, medicine, etc. in the healthcare systems can be constantly tracked and monitored. As an example, a patient's heart rate can be sensed frequently, and data be transmitted to the physician for assessment and monitoring using the personal computing and handheld devices like mobile-phone, wearables, etc.[15].

The following contains discussion for IoT cloud-based big-data supported context-aware systems, middleware or framework proposed in the last few years proving the latest development and research capabilities in this field.

Hydra [16] is an IoT based middleware, which is designed to deliver solutions using distributed architecture for heterogeneous wireless physical devices and sensors into the ambient intelligence environment. Its prime objective was to provide connection to embedded devices to the applications. The special environment initiated by Hydra contains applications for smart homes, smart hospital, and smart farm as use cases. These applications are built on service-oriented architecture, which are well equipped with powerful reasoning toward various context sources. Hydra works on two levels i.e. data acquisition component, which is responsible for the sensory connection and data retrieval. On the second level, i.e. modeling and reasoning level, it uses hybrid modeling techniques such as key-value, ontology-based and object-oriented methods.

Feel@Home [17] a context management framework interacting with several domains such as controlling wireless sensor networks through an Internet for an IoT based smart environments like offices/homes. This framework provides and demonstrate baseline paradigm, which can be utilized in any of the smart environment by accessing sensors and actuators using a gateway, and fusing them to support diverse application domains.

Semantic Web-based Context Management (SeCoMan) [18] provides user's, a privacy-preserved solution by providing opportunities to develop context-aware smart applications in the IoT. The modules comprising SeCoMan are coupled in a layered structured architecture, which includes the application layer for desired services for users, context management layer for supporting context-aware applications, and plug-in sensors used as location-based independent context sources. SeCoMan utilizes semantic-oriented IoT vision where semantic technologies based on ontology plays a key role to model the user's indoor locations and objects to obtain valuable knowledge under defined context-aware policies.

Authors of [19] presented a novel Cloud-oriented Context-Aware Middleware in Ambient Assisted Living (CoCaMAAL), which strives to ease the daily lives of people suffering from some chronic medical disabilities. This framework, by utilizing several components, enhances monitoring and data aggregation capacity for biomedical sensors. It is comprised of five main cloud-oriented components: (1) Ambient Assisted Living (AAL) systems based on Body Sensor Network; (2) context aggregator and Google supported context providers cloud manipulating, recognizing and classifying activity and healthcare sensory data. into context based on ontology; (3) Service providers cloud capable of handling healthcare context-aware and healthcare emergency detection services; (4) context-aware middleware (CaM) cloud coordinating and managing context functions among context data store and service providers cloud and (5) context data visualization cloud. All these components provide methods for processing heterogeneous data such as vital signs, daily life activities, and location log to obtain high-level contexts using generalized medical rules in the ontology. This work was further extended by addressing personalized knowledge discovery using Big Data for Context-aware Monitoring (BDCaM) framework, which aimed to provide more useful information for context-aware decision by discovering personalized knowledge from patient data. It provided real-time personalized health-care services using big data application.

FlexRFID [20] a multi-layered flexible and highly configurable middleware solution to facilitate context-aware applications development by integrating heterogeneous IoT devices like RFID and WSN. The experimental results were evaluated on healthcare scenario and adopted cloud-based policies to process preferred data based on application rules and preferences. It supported and claimed provision of the patient's safety by speeding up medical treatment and saving time, cost with improved operational efficiency.

Internet of Things based Physical Activity Monitoring (PAMIoT) [21] works on top of a variety of biosensors, and monitors the physical activity of an ordinary person by enabling IoT cloud-based services for the pulse oximeter, ECG, and foot counts.

Context Awareness for the Internet of Things (CA4IOT) [15], a sensing architectural middleware service to help users by automating the task of selection and actuation of the sensors as per requirements. It works around operations on layered architecture by allowing for fine-tuning to address the disparity in contextual limitations. These phases allow automated filtering, management and reasoning in the realm of sensor data acquisition, thus creating meaningful information from raw sensory data.

Another prominent work in this regard provides an insight with an aim to underpin the vision of the platform comprising of Sensor Platform for Healthcare in a Residential Environment (SPHERE) project[22]. The proposed framework fuses complementary sensor data, which leads to the generation of rich datasets for monitoring and managing various health conditions. It works around integrated multimodality sensing system infrastructure.

Multi-level Context-aware Framework (mlCAF) [23] mechanizes automatic identification of the user's contexts across the cross-domain environments in ubiquitous and pervasive computing systems. It helps and demonstrates ontology modeling of contexts at different levels by preserving the user's behavior. These obtained contexts are monitored continuously, which leads to help in controlling lifestyle associated chronic diseases such as diabetes. This framework uses IoT enabled sensory raw data to model heterogeneous contexts for multiple domains such as physical activity, nutrition and clinical for context-awareness. These contexts are fused together semantically to arbitrate richer behavioral contexts using an open-source ontology.

2. Lifelog Monitoring

Routine is flourished with time to become habits and these habits reflect the futuristic trends of diseases[24]. But the indication for the prolonged unhealthy activities and consumption of unhealthy nutrient in diet is quite supportive for chronic disease management. It is quite necessary to monitor the wellness attributes like diet and physical activities for taking proactive remedies. Intelligent and sophisticated tools and techniques are required to focus on self-monitoring. For this we need to put stronger emphasis on balanced diet, non-sedentary lifestyle and daily physical activities. Wellness is a very concerned area regarding the prevention of diseases to improve health and lower the costs for healthcare. Wellness monitoring and management is to bring in the right set of healthcare providers at the right time on the continuum of care, starting from prevention to curing for assuring therapy adherence. Wearable and smartphone based wellness applications are available in market to provide health recommendation based on user preferences, lifelog, personalized goals, calories consume, and calories burn. The commercial wellness applications consist of sensors to detect and record the activities and desktop or mobile application for the visualization of the lifelog and recommendation. The applications of physical wellness can be divided into 3 main categories

2.1. Fitness Monitoring

The applications related to user fitness deal with physical activity and helps the user to plan goal for calories burning and distance covered. Usually these applications provide proper time to rest and restore the body for the attaining the goal in a week. These applications train and enhance the level of physical activities by monitoring the status of the user gradually.

Johnson & Johnson Official 7-Minute Workout[1)] is an application that is designed to maximize the activities results and make every minute count. It has helped more than 1.4 million people get fit in just seven minutes each day by providing more than one thousand exercises. It provides support to customize the workout routine and creates healthy competitive environment among the friends by sharing the progress. The application has multiple intensity levels of workout to enhance the ability of the customers gradually from beginner level to professional level. It has the features of workout assessment and setting up the personalized challenge level.

1 https://7minuteworkout.jnj.com/

Blogilates[2] is an application that provides the YouTube videos for the training of fitness exercises. It is one of the leading channel on YouTube of female fitness. It has over 300 million video views and 3 million subscribers. With the Blogilates app, customers access to their daily recommended workouts with exclusive workout videos, and recipes. The application provides sorting and searching facilities by series and by body parts. It creates competitive environment through sharing the progress pictures with the Blogilates community. The workout calendars and recipes support the customers for setting up the personalized goal and challenges.

Cody[3] is an application that provides the training and tracking with the help of world-class coaches through smartphone with the help of mobile application. It is a video-training application that supports the customers to remain consistent and know about their progress. The guiding feature of the application makes the users to learn from world-class coaches the new skills and reach new heights. The lifelog tracker tracks the progress and maintain it. Users can access videos offline and at any time.

Moves[4] is an application to track the lifelog activities. It maps the daily activities and different aspects to monitor the health progress of the customers. The application maintains the log of the activities, their duration, distance, number of steps, and calories burned. The key features include the gym training and calculates the calorie burn for activities and visualized daily idle calorie burn.

Nike+ Training Club is a very comprehensive fitness application. The application is designed by professionals. It has the capabilities to provide customizable training workout for the customers[7]. The customers are provided the capabilities of sharing and comparing the progress with their relatives and companions to build a sense of challenge and competition. This app creates a personalized training plan based on our needs, then continually adapts it as you go.

2 https://www.blogilates.com

3 https://www.codyapp.com

4 https://moves-app.com/

2.2. Nutrition Monitoring

Multiple researches have shown that specific unhealthy behaviors, including smoking, higher alcohol intake and, unbalanced diets are associated with an increased risk chronic diseases and premature mortality[25]. The health status of the population can simply be measured from mortality rate, where unhealthy behaviors have been associated with a higher risk of lifestyle diseases[26]. Diet is an influential health agent and bad diet pattern is among the prominent causes of premature death and chronic disease[27]. Current wellness applications recognize the user activities, log them and represent in an interactive graphical manner. In addition to present the user's activities logs, a range of applications also provides the impact of co-related activities on health and recommendations. It is more beneficial to indicate the unhealthy nutritional habits to the healthcare stakeholders which may help them to avoid their bad impact. A lot of nutrition applications are available which monitors and give recommendation to the customers.

DietPoint[5] is an application that gives control over the diet to reduce and maintain a healthy weight. It consists of more than 130 diet plans which help to reduce and maintain the desired weight. Customers have flexibility to acquire the best suitable diet plan according to their requirements and lifestyle. The application supports for the body mass index calculator along with dedicated shopping list with respect to the desired diet plan. It monitors and tracks the weight and provides the weight-loss estimation with respect to the plan.

Fooducate[6] is an application which supports beyond a simple calorie counters. It records the grocery, sleep, mood and hunger. It informs about the healthiness of different food ingredients and nutrients. It analyzes the recorded data and information and gives recommendation that help users to attain the desired health and fitness status. The app track what the user eats, what activities do and how much calories burnt. According the user profile information, it gives food and activity recommendations.

5 www.dietpointed.com

6 http://www.fooducate.com/

Lifesum[7] is an application that helps the customers to adapt healthy dietary behavior in a progressive manner based on personalized plans targeting the user goal. It notifies the users regarding hydration and food intake based on the personalized goal. It gives feedback on food intake and visualization of customer progress. The application provides customers with personalized roadmap of how to change diet for the better weight with exercise. It helps the customers for the selection of diets and give feedback to improve the quality diet, nutrition, and exercise.

Lose It![8] Is another application which helps to monitor the nutrition to manage calories goal. It keeps track of calories input through pictures of the food. The application has support to integrate with other multiple wellness applications to track the fitness level. It provides personalized plan for calories intake and weight loss. It tracks calories through food pictures and integrate with other fitness applications for seamless activity and biometric tracking.

For Dietary management applications the most important requirement is the database of foods and their nutrition. MyFitnessPal[9] is one of the very famous application in the domain of health and fitness. It has a support of database which contains more than 6 million food items. It can track almost everything that the user eats. It recommends the best diet which helps to meet the personalized goal. First the user sets the goal, then the app tracks the food intake and user activity and recommend some food and activities to achieve that goal.

My Diet Coach App[10] motivates its users to lose weight, achieve goal and remain healthy. It indicates the customers regarding proper hydration, preparing vegetable food and eating properly in slow speed. It maintains diet diary and calories counter along with points and rewards based influence.

2.3. Healthcare Monitoring

From counting daily calories intake to monitoring blood sugar levels applications are monitoring everything for the improvement of human life. The healthcare applications have increased in the past few years due to advancement in the development of ubiquitous computing with IoT support. Medical instruments are supported with computer software and these applications manage data from the device and provide visual analytics to the stakeholders on their demand[11]. Currently, the smartphones market has increased the accessibility of the healthcare management applications. These applications offer basic as well as advanced features like blood glucose trends, information sharing with stakeholders, virtual logbooks, informative knowledge, and more.

Endomondo[12] is an application to provide support for Cardio exercise. Users usually lost their interest in same and repetitive workouts. The application provides customers the flexibility to set their goals for running, walking, biking or riding. It provides a virtual trainer who monitors the activities and guide for the goals achievement and improvement in pace. Users can compete with themselves or their friends. It tracks workouts, provides audio feedback along the way and offers guidance on how to reach your goal. It is integrated with a wide range of smart watches and sensors to enhance the user experience and provide user with more comprehensive workout data such as heart rate.

Samsung Health[13] is a personal trainer like application that helps users to attain the daily goal in term of calories intake and burn. The application is enriched with multiple training programs to support customers to maintain healthy lifestyle and fitness level. The application uses different kinds of tracker to capture a wide range of physical activities in different environments.

Apple Health Kit[14] is a comprehensive application for both wellness and health. It takes care of fitness attributes as well as health vital signs. The application provides support for automatic tracking of physical activities along with manual entries of medical records like BMI, body

7 https://lifesum.com/
8 https://www.loseit.com/
9 https://www.myfitnesspal.com/
10 https://www.mydietcoachapp.com

11 http://static.diabetesselfmanagement.com/pdfs/DSM0312_011.pdf
12 https://www.endomondo.com/features
13 https://www.sammobile.com/apk/samsung-health/s-health-5-7-0-0025/
14 https://developer.apple.com/reference/healthkit

temperature and other vital signs. The application is designed in a way such that the data collected from it can be shared with clinicians. Health kit is compatible with more than 100 different applications with direct support to Apple Watch as a wearable.

Withings Health Mate[15] is focused to track the activities through accelerometer of a smart phone. Users can share and compete with their friends regarding the walk per week through application available in their smartphone which tracks the activity. Virtual rewards and achievement badges concept is used to motivate the users for physical activities. Application can analyze the activity level against usual. It informs for unusual level of activities to the user.

Ⅲ. Big Data Focused Systems

1. Visualization and Analytics

Nowadays with the advancement of healthcare industry, it generates a large amount of data in term of records, compliance, and regulatory requirements. As healthcare contains different type and volume of data; therefore, big data is an essential implementation. The data normally consist of clinical data, wellness data and patient records. These data help in identifying the disease patterns and quality of life by lifestyle; furthermore, it helps in improving prognosis and recommendation generation by physicians and health experts. We have further classified this section in two categories:

1.1. Big Data Core for Analytical Healthcare

Variety of big data analytical methods are used in medical domain to improve the quality of care. The aim of healthcare analytics is to understand the underlying semantics of medical data. Similarly, medical imaging techniques are employed to visualize interior of body for healthcare analysis where physicians are required to correctly interpret these images to provide effective medication. Some of the medical modalities are radiography, magnetic resonance Imaging (MRI), ultrasound, electrography and tomography. Many techniques have been proposed for analyzing medical health data; however, most of these techniques expose a common limitation of handling large datasets. To overcome this limitation Hadoop-based resolutions comes into play, which intern employs a technique called map-reduce[28][29].

Map-reduce is well known for its scalability for large datasets over large number of servers in a Hadoop cluster [30], and thus having an array of real-world applications in diverse domains[31][32]. Every technique comes with pros and cons and same goes for map-reduce. Map-reduce based techniques show difficulty when faced with I/O intensive tasks[30]. In one of the research [30] map-reduce is employed in medical domain for variety of experiments on images like finding optimal hyper-parameters for lung texture classification by adopting support vector machine (SVM), content-based image recognition (CBIR) on medical image indexing and analysis of wavelet, that greatly reduced the overall analysis and processing time for these large set of medical images.

The construction of hair-trigger methods for applications where the existence of disease is considered the most critical for successful emergency treatment right away, is compelling as in such cases the aim is to exploit the analysis of these methods in the emergency treatment[33]. Thus, the processing time expediency of such methods is essential. Similarly, to achieve better scalability the computational complexity of the methods (algorithms) needs to be reduced, likewise the increase in computational complexity of the algorithms kills the scalability on the other hand. For instance, to achieve scalability in information retrieval Tsymbal et al. proposed an approach for clinical decision support systems (CDSS) that utilizes discriminative distance learning along with extremely lower computation complexity and hence achieved better scalability[34].

Healthcare being the most critical domain requires development of scalable and performance efficient methods with no accuracy loss. Map-reduce splits the data into smaller parts and then maps each part

15 https://www.withings.com

to separate reducers for processing; therefore, the determination of dependencies among the data can help improving the accuracy. As an example, a hybrid machine learning approach has been employed in [35] where the classification among the schizophrenia patients and healthy controls have been performed by utilizing functional magnetic resonance imaging (fMRI) with a nucleotide polymorphism (SNP) input[35]. An accuracy of 87% has been claimed by the authors which would not be possible if either of fMRI or SNP have been used in an isolation.

An experiment on the comparison of organ segmentation methods for big data have been performed by del Toro and Muller. They designed a method that encompasses atlas probabilistic information and local contrast of medical images[36]. An overall 33% of improvement in the accuracy has been claimed by the authors by utilizing atlas probabilistic information with local contrast of the medical images.

Similar techniques have been employed in the past research on designing clinical decision support systems (CDSS) that incorporates interdependent patient data. For instance, an approach for clinical decision support system is proposed by Chen et al. [37] that intelligently combined various related information associated with patient like demographic, medical, and computerized axial tomography (CT) scan. This approach provides prediction for the level of intracranial pressure (ICP). The developed CDSS system facilitates the physician while taking decisions to provide treatment for the patients that suffer from traumatic brain injury (TBI)[37].

New innovative techniques of big data analytics in healthcare are being introduced in research community. Seton Healthcare are constantly exploiting the results of predictive analytics to minimize the high-cost congestive heart failure readmissions[38]. Similarly, “MediSys” has been proposed by European commission where it acts as a scanning tool for bioterrorist activity detection and close observation of communicable diseases. On daily basis, target news can be caught by “MediSys” from vast amount of internet articles generated by European media monitor. This target news can be sent by the “MediSys” to the required patient by means of email and SMS. Numerous conventional problems in healthcare domain have been resolved by employing big data analysis tools and techniques[39]. Advancement in quality of care, rise in effectiveness, and reduction in number of readmissions in the healthcare centers are some of the examples achieved via big data analytics.

1.2. Healthcare Information Visualization

An earliest research on data visualization in healthcare domain was published by Powsner and Tufte[40]. The proposed methodology facilitates physicians by providing a holistic view of the patient status by visualizing patient’s health record. Similar research has been published in[41]. In this research, lifelines technique has been developed that utilizes various graphical elements to draw a comprehensive view of patient’s health status. Another effective method for patient health data visualization called Knowledge-based Navigation of Abstractions for Visualization and Explanation (KNAVE) is proposed by Shahar and Cheng[42][43]. KNAVE provides a convenient way for visualizing time-oriented EHR data. In this technique graphs and charts have been employed for better understandability of information contained in the data.

A unique approach for visualization of EHR data has been proposed by Gotz et al.[44]. The proposed technique is called Dynamic Icons (DICON). In this approach EHR data of various patients have been employed to create clusters of similar patients. DICON provides a comfortable way of interaction with patient clusters and perhaps a convenient way for healthcare practitioners to understand similarity among various patients.

Another technique proposed by Joshi and Szolovits [45] utilizes radial starburst approach on very large dataset (EHR) to present the complexity of data depicted on a 100-dimensional space. To reduce the complexity of data, machine learning technique has been used to group the clusters of similar patients. Each cluster was defined by eight psychological foci. The proposed decision-making tool can visualize a very large and complex EHR data.

2. Data-driven Knowledge Generation

In the current era, knowledge generation is one of the key research area as new knowledge can play an important role for better analysis in healthcare or wellness domain. Normally, in big data domain, three concepts of data (i.e. structured, semi-structured and unstructured data) are used, where each concept has its own methods to extract useful information from a given data. For providing better healthcare services, it is crucial and expensive to extract knowledge through manual means. For knowledge generation process, more systematic methodologies have been devised, which can be termed as the data-driven knowledge generation. These methodologies can improve the performance of a domain expert and can reduce the cost of creating new knowledge from data.

In big data domain, a term 'Big Data Mining' is used for extracting hidden information from a huge volume of data to generate the knowledge. Big data mining is a promising research area and an emerging trend in all domains. Three principals are mentioned in [46] for extracting knowledge from big data, which are (1) capability of a system to perform visualization, statistical analysis, machine learning, and data mining tasks, (2) support of multiple storage mechanism, and (3) ability to access the results with ease for understandability.

For extracting hidden information from a large volume of data, the DDKAT, NIMBLE, Apache Mahout, BC-PDM, SAMOA, DMCF, and others are the state-of-the-art big data mining tools/platforms/frameworks. Followings are the brief description of each tool/platform/framework.

For wellness domain, a Data-Driven Knowledge Acquisition Tool (DDKAT) for the Mining Minds platform has been developed [47], which generates and shares the production rules by utilizing the user profile and lifelog data stored in the big data storage[48]. The DDKAT helps to evolve the wellness knowledge for providing a better quality of service[48]. This tool covers all major aspects of data mining process and having the capability of an end-to-end knowledge engineering process. For knowledge generation, it includes data selection, data pre-processing, model learning, model translation, and rule conformance phases[48]. In data selection phase, big data is involved, where a domain expert loads the schema of a historical data maintained in the distributed big data storage to select the required parameters. On the backend, a Passive Data Reader (PDR) component has been designed [13], which transfers the schema to the domain experts and after parameters' selection, it executes the domain expert's query over big data storage and retrieves the required parameters' data from Hadoop Distributed File System (HDFS). All communication of the PDR is handled in a JSON format. In data pre-processing phase, missing values, outliers, and data discretization tasks are performed. For model learning phase, only decision tree classifiers are used. In next phase, classification models are translated into an XML format to extract the executable rules, which are conformed from a domain expert for verification purposes[48].

NIMBLE [49] is a domain-specific portable infrastructure, which parallelizes the machine learning and data mining algorithms (ML-DM) for rapid computation. For parallelization, Hadoop framework has been used in this study. NIMBLE has an ability to process multiple data formats and can support MPI and X10 runtimes. The IBM Corporation has implemented this infrastructure. Ghoting et al. [49] demonstrated the k-Nearest Neighbors, k-Means Clustering, Pattern Growth-based Frequent Item set Mining, Random Decision Trees, and RBRP-based Outlier Detection algorithms to evaluate the expressiveness and performance of the NIMBLE.

Apache Mahout [50] is open source java project launched by Apache Software Foundation. It provides support of scalable data mining algorithms for clustering, categorization, evolutionary programming, and collaborative filtering (CF) tasks. Mahout's primary features are: (1) Taste CF, (2) supports Canopy, k-Means, fuzzy k-Means, Mean-Shift, and Dirichlet, clustering algorithm implementation with MapReduce model, and (3) Distributed and Complementary Naive Bayes classifications' implementations. Similarly, Scalable Advanced Massive Online Analysis (SAMOA) [51] is a platform for mining big data streams using distributed algorithms. It is written in Java language and can run on various distributed stream processing engines such as Samza Storm, and S4. SAMOA supports the common machine learning tasks such as regression, clustering, and classification.

Authors of [52] proposed a platform to provide personalized ubiquitous life care (u-life care) services. The proposed platform processes the structured as well as unstructured data, extracts knowledge while exploiting big storage technology for massive storage of data, and provides a consolidated service as well as analytics to assist in decision making. It supports data acquisition and management, big data storage and processing, data wrangling, learning models, construction of knowledge bases, and finally u-life care services API. Furthermore, this platform can assist the individuals to visualize their personal behavior patterns to manage their daily-life activities.

In healthcare domain, researchers of [53] described a generic prediction framework for malaria treatment outcomes using big data analytics, which was employed under the umbrella of the Hadoop MapReduce framework. The author used Waikato Environment for Knowledge Analysis (WEKA) to generate the best ten rules. In this framework, data were collected from hospitals as well as from individual patients and then stored at a central location for performing big data analytics. After selecting a mining technique, an appropriate mining algorithm is applied to generate the best rules.

Authors of [54] described the process of making knowledge discovery service as scalable and designed the Data Mining Cloud Framework (DMCF) to develop and execute the distributed data analytics applications for services purposes. Author has used datasets, data mining algorithms, analysis tools, and knowledge models as single services and then combined them through an interface to execute it on cloud. The DMCF framework allows users to implement single-task, parameter-sweeping, and workflow-based applications. The proposed framework also supports the composition as well as the execution of the workflow-based data-driven knowledge discovery applications. The knowledge discovery workflows help to

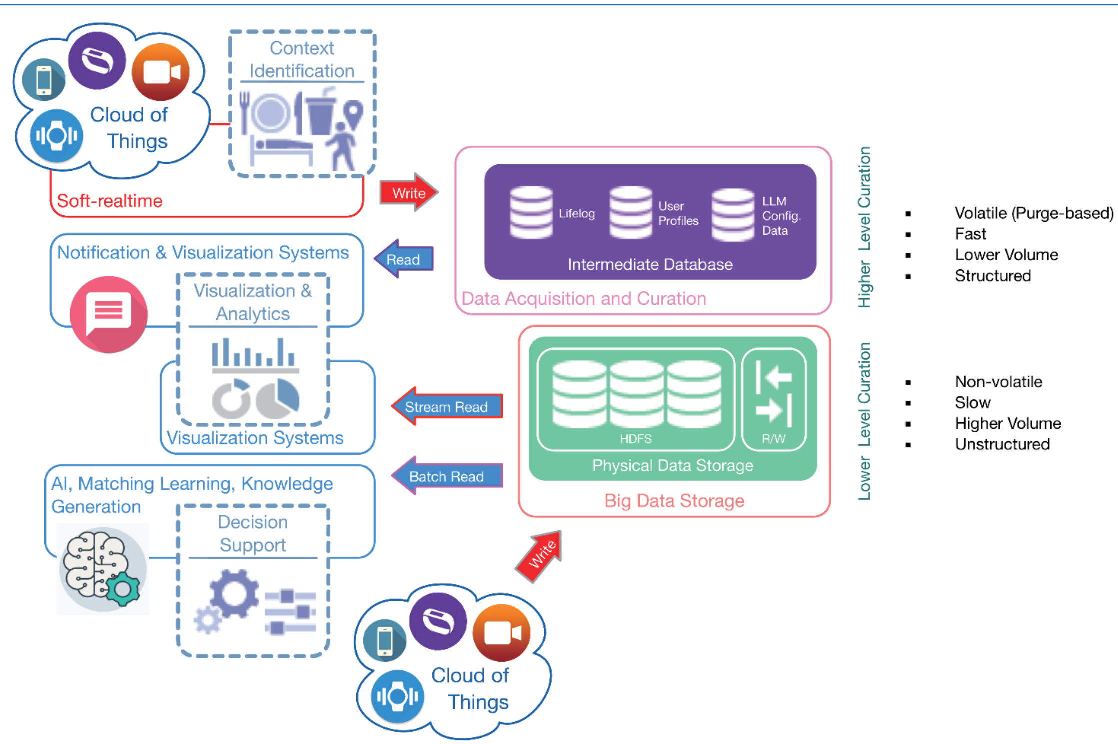


Fig. 4. Multi-level Data Curation

confirm the real-world experiments and provide insights. To discover hidden knowledge from a dataset, author illustrated a parallel classification workflow.

Authors of [55] designed and implemented a framework to discover hidden knowledge from a large-scale healthcare big data using various data preprocessing and data mining techniques. For data preprocessing, data cleaning to remove inconsistent and missing values, data integration, transformation, and data reduction tasks were involved. Similarly, for data mining techniques, classification algorithms are considered for performing predictive tasks, while cluster analysis and association rules are for descriptive tasks. Finally, data mining visualization techniques are used to determine the hidden patterns from large-scale databases and then to extract visual patterns for end users in less time. The proposed framework can assist the healthcare practitioners in medical diagnosis. In addition, authors described a role of big data in the medical application area. For case study purposes, the diabetic patients admitted to a hospital were considered in this study.

IV. Data Curation Framework

This section provides an overview of the proposed methodology of Data Curation Framework (DCF). Special focus is provided on the curation strategy implemented by DCF. Technical and implementation details of this framework are published in[13].

DCF is a platform layer for other health and wellness systems where real-time data acquisition from IoT, its curation as a lifelog, and monitoring for anomalies is essential. As illustrated in the architecture 〈Fig. 1〉, DCF consists of two internal layers, i.e., Data Acquisition and Curation, and Big Data Storage. Data Acquisition and Curation subcomponents obtain the sensory data from IoT devices. Current implementations of Cloud-centric IoT heavily depend on device's hardware and its embedded system. Thus, communication occurs with the protocols defined by the APIs of device manufacturer. Therefore, configurator part of this subcomponent overrides the communication between the cloud and the sensory device.

Data Synchronization subcomponent acquires the raw sensory data from multimodal data sources, both in a soft-real and offline. This data is synchronized based upon the user identification and the time stamp of the data generation, and subsequently, queued for the context determination.

The subcomponent of Lifelog Representation curates the user context by mapping it to a time-based log registering the detected human activities and behaviors. These lifelog instances are analyzed by monitoring subcomponent called Lifelog Monitor (LLM). It is responsible for performing customized situation-based (e.g., time-based, physical activity-based, nutrition-based) monitoring of user context available in the lifelog, cross-linked with the user profiles. LLM provides a notification-based response with the help of a publish-subscribe model.

The Big Data Storage component is responsible for providing permanent and distributed big data persistence to the raw data acquisitioned from IoT. Behavior of this component is non-volatile persistence, as no update or delete operations are performed on the stored data. Big Data Storage also provides means to response in an online or offline manner. For online response, the Stream Data Reader subcomponent is used. It provides a continuous stream of sensory data for extended data operations, including visualization and predictive analytics. For an offline response, the Batch Data Reader subcomponent is used. It provides a batch-based response that can be effectively used for training machine learning-based models and can provide data insights to experts for rule generation.

1. Multi-level Big Data Curation

The asset of DCF is its persistence of a user's raw sensory data (from IoT devices) with the associated context as a lifelog. DCF defines two levels of curation on the data acquired 〈Fig. 4〉. The first level is a higher-level abstraction referred to as an Intermediate Database hosted within a relational database (RDBMS). This database hosts three types of data, i.e., the user lifelog that represents user context over a period, the user profiles, and the knowledge base consisting of rules for monitoring.

The second level is a lower-level abstraction referred to as big data storage; it is hosted over Hadoop-based big data cluster. This storage also hosts three types of data; however, the granularity of data is at a finer scale. For example, this storage provides permanent persistence to all the raw sensory data acquired from the multimodal data sources; it maintains user-invoked backup of large-sized multimedia content, such as video data captured from a 3D camera and periodic backups of user lifelogs with associated user profiles.

V. Conclusion

In this paper, we reviewed the state-of-the-art tools, techniques, applications, and methodologies designed to cater the needs of heath and wellness applications with large-scale usage and datasets. With the inclusion of IoT as the source of origination of big data, Cloud-centric IoT is achieving the goal of a centralized hub of health and wellness data for users with control over access depending upon the service scenarios and use cases.

Acknowledgment

This work was supported by the Korea Research Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2016H1D3A1938039), and this research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-0-01629) supervised by the IITP(Institute for Information & communications Technology Promotion)

References

[1] Banos, O.; Amin, M.B.; Khan, W.A.; Afzal, M.; Hussain, M.; Kang, B.H.; Lee, S.Y. The Mining Minds Digital Health and Wellness Framework. BioMedical Engineering OnLine 2016.

[2] Hood, L.; Flores, M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. New Biotechnology 2012, 29(6), 613 - 624.

[3] S. Marusic J. Gubbi, R. Buyya and M. Palaniswami. 2013. Internet of Things(IoT) - A vision, architecture elements, and future directions. Future Generation Comp. Syst. 29, 7 (2013), 1645 - 1660.

[4] Dake Liu Bo Hu Shanzhi Chen, Hui Xu and Hucheng Wang. 2014. A Vision of IoT: Applications, Challenges, and Opportunities With China Perspective. IEEE Internet Things J. 1, 4 (2014), 349 - 359.

[5] A. Pitsillides A. Kamilaris and V. Trifa. 2011. The Smart Home meets the Web of Things. IJAHUC 7, 3 (2011), 145.

[6] B.KantarciandH.T.Mouftah.2015.Sensing services in cloud-centric Internet of Things: A survey, taxonomy and challenges. In 2015 IEEE International Conference on Communication Workshop (ICCW). 1865 - 1870. https://doi.org/10.1109/ICCW. 2015.7247452

[7] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. Foundations and Trends in Information Retrieval 8, 1 (2014), 1 - 125. https://doi.org/10.1561/1500000033

[8] Zini, F.; Reinstadler, M.; Ricci, F. Life-logs Aggregation for Quality of Life Monitoring. Proceedings of the 5th International Conference on Digital Health 2015; ACM: New York, NY, USA, 2015; DH '15, pp. 131 - 132.

[9] Powell, A.C.; Landman, A.B.; Bates, D.W. In search of a few good apps. The Journal of the American Medical Association 2014, 311(18), 1851 - 1852.

[10] Banos, O.; Villalonga, C.; Damas, M.; Gloesekoetter, P.; Pomares, H.; Rojas, I. Physiodroid: Combining wearable health sensors and mobile devices for a ubiquitous, continuous, and personal monitoring. The Scientific World Journal 2014, 2014(490824), 1 - 11.

[11] Oresko, J.J.; Jin, Z.; Cheng, J.; Huang, S.; Sun, Y.; Duschl, H.; Cheng, A.C. A Wearable Smartphone-Based Platform for Real-Time Cardiovascular Disease Detection Via Electrocardiogram Processing. IEEE Transactions on Information Technology in Biomedicine 2010, 14, 734 - 740.

[12] Zini, F.; Reinstadler, M.; Ricci, F. Life-logs Aggregation for Quality of Life Monitoring. Proceedings of the 5th International Conference on Digital Health 2015; ACM: New York, NY, USA, 2015; DH '15, pp. 131 - 132.

[13] Muhammad Bilal Amin, Oresti Banos, Wajahat Ali Khan, Hafiz Syed Muham- mad Bilal, Jinhyuk Gong, Dinh-Mao Bui, Soung Ho Cho, Shujaat Hussain, Taqdir Ali, Usman Akhtar, Tae Choong Chung, and Sungyoung Lee. 2016. On Curating Multimodal Sensory Data for Health and Wellness Platforms. Sensors 16, 7 (2016).

[14] M. C. Domingo, "An overview of the internet of things for people with disabilities," J. Netw. Comput. Appl., vol. 35, no. 2, pp. 584 - 596, 2012.

[15] Perera, C.; Zaslavsky, A.; Christen, P.; Georgakopoulos, D. Context aware computing for the internet of things: A survey. IEEE Commun. Surv. Tutor. 2014, 16, 414 - 454.

[16] A. Badii, M. Crouch, and C. Lallah, "A context-awareness framework for intelligent networked embedded systems," in Proc. 3rd Int. Conf. CENTRIC Adv. Human-Oriented Pers. Mech., Technol. Serv., Aug. 2010, pp. 105 - 110.

[17] B. Guo, L. Sun, and D. Zhang, "The architecture design of a crossdomain context management system," in Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops, Mar. 2010, pp. 499 - 504.

[18] Huertas Celdran, A.; Garcia Clemente, F.J.; Gil Perez, M.; Martinez Perez, G. SeCoMan: A Semantic-Aware Policy Framework for Developing Privacy-Preserving and Context-Aware Smart Applications. IEEE Syst. J. 2013, doi:10.1109/JSYST.2013.2297707.

[19] Forkan, A.R.M.; Khalil, I.; Ibaida, A.; Tari, Z. BDCaM: Big Data for Context-aware Monitoring—A Personalized Knowledge Discovery Framework for Assisted Healthcare. IEEE Trans. Cloud Comput. 2015, 99, doi:10.1109/TCC.2015.2440269.

[20] El Khaddar, M.A., Chraibi, M., Harroud, H., Boulmalf, M., Elkoutbi, M. and Maach, A., 2015. A policy-based middleware for context-aware pervasive computing. International Journal of Pervasive Computing and Communications, 11(1), pp.43-68.

[21] Ray, P.P., 2014. Internet of things based physical activity monitoring (PAMIoT): an architectural framework to monitor human physical activity. Proceeding of IEEE CALCON, Kolkata, pp.32-34.

[22] Zhu, N., Diethe, T., Camplani, M., Tao, L., Burrows, A., Twomey, N., Kaleshi, D., Mirmehdi, M., Flach, P. and Craddock, I., 2015. Bridging e-health and the internet of things: The sphere project. IEEE Intelligent Systems, 30(4), pp.39-46.

[23] Razzaq, M.A., Villalonga, C., Lee, S., Akhtar, U., Ali, M., Kim, E.S., Khattak, A.M., Seung, H., Hur, T., Bang, J. and Kim, D., 2017. mlCAF: Multi-Level Cross-Domain Semantic Context Fusioning for Behavior Identification. Sensors, 17(10), p.2433.

[24] Tremblay, Mark Stephen, et al. "Physiological and health implications of a sedentary lifestyle." Applied Physiology, Nutrition, and Metabolism 35.6 (2010): 725-740.

[25] Kvaavik, Elisabeth, et al. "Influence of individual and combined health behaviors on total and cause-specific mortality in men and women: the United Kingdom health and lifestyle survey." Archives of internal medicine170.8 (2010): 711-718

[26] Petersen, Kristina EN, et al. "The combined impact of adherence to five lifestyle factors on all-cause, cancer and cardiovascular mortality: a prospective cohort study among Danish men and women." British Journal of Nutrition 113.05 (2015): 849-858

[27] Katz, D. L., and S. Meller. "Can we say what diet is best for health?." Annual Review of Public Health 35 (2014): 83-103.

[28] Shvachko, Konstantin, et al. "The hadoop distributed file system." Mass storage systems and technologies (MSST), 2010 IEEE 26th symposium on. IEEE, 2010.

[29] Sobhy, Dalia, Yasser El-Sonbaty, and Mohamad

Abou Elnasr. "MedCloud: healthcare cloud computing system." Internet Technology And Secured Transactions, 2012 International Conference for. IEEE, 2012.
[30] Markonis, Dimitrios, et al. "Using MapReduce for large-scale medical image analysis." Healthcare Informatics, Imaging and Systems Biology (HISB), 2012 IEEE Second International Conference on. IEEE, 2012.
[31] Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." Communications of the ACM 51.1 (2008): 107-113.
[32] Li, Wen-Syan, et al. "Xbase: cloud-enabled information appliance for healthcare." EDBT. 2010.
[33] Shackelford, Karen. "System & method for delineation and quantification of fluid accumulation in efast trauma ultrasound images." U.S. Patent Application No. 14/167,448.
[34] Tsymbal, Alexey, et al. "Towards cloud-based image-integrated similarity search in big data." Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on. IEEE, 2014.
[35] Yang, Honghui, et al. "A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia." Frontiers in human neuroscience 4 (2010).
[36] del Toro, Oscar Alfonso Jimenez, and Henning Muller. "Multi atlas-based segmentation with data driven refinement." Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on. IEEE, 2014.
[37] Chen, Wenan, et al. "Intracranial pressure level prediction in traumatic brain injury by extracting features from multiple sources and using machine learning methods." Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on. IEEE, 2010.
[38] IBM. Reducing readmissions to improve care [Internet]. Armonk (NY): IBM; c2014 [cited at 2014 Oct 20]. Available from: http://www-01.ibm.com/software/sg/ industry/healthcare/pdf/setonCaseStudy.pdf.
[39] European Commission. Medical intelligence in Europe [Internet]. Brussel: European Commission; c2014 [cited at 2014 Oct 20]. Available from:http://ec.europa.eu/health/preparedness_response/generic_preparedness/ planning/medical_intelligence_en.htm
[40] Powsner, Seth M., and Edward R. Tufte. "Graphical summary of patient status." The Lancet 344.8919 (1994): 386-389.
[41] Plaisant, Catherine, et al. "LifeLines: visualizing personal histories." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 1996.
[42] Shahar, Yuval, and Cleve Cheng. "Intelligent visualization and exploration of time-oriented clinical data." Systems Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on. IEEE, 1999.
[43] Shahar, Yuval, and Cleve Cheng. "Model-Based Visualization of Temporal Abstractions." Computational Intelligence 16.2 (2000): 279-306.
[44] Gotz, David, et al. "Visual cluster analysis in support of clinical decision intelligence." AMIA Annual Symposium Proceedings. Vol. 2011. American Medical Informatics Association, 2011.
[45] Joshi, Rohit, and Peter Szolovits. "Prognostic physiology: modeling patient severity in intensive care units using radial domain folding." AMIA Annual Symposium Proceedings. Vol. 2012. American Medical Informatics Association, 2012.
[46] Alotaibi, N.M. and Abdullah, M.A., 2017. Big data mining: A classification perspective, pp. 687-695.
[47] M. Ali, "Ddkat - data driven knowledge acquisition tool, version 1.0," Available online: https://github.com/ubiquitous-computing-lab/Mining-Minds/tree/master/knowledge-curation-layer/DDKAT, 2017, accessed:2017-12-29.
[48] Ali, M., Hussain, M., Le, S. and Kang, B.H., 2016. SaKEM: A Semi-automatic Knowledge engineering methodology for building rule-based knowledgebase. In 16th International Symposium on Perception, Action, and Cognitive Systems (PACS2016) (pp. 63-64).

[49] Ghoting, A., Kambadur, P., Pednault, E. and Kannan, R., 2011, August. NIMBLE: a toolkit for the implementation of parallel data mining and machine learning algorithms on mapreduce. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 334-342). ACM.

[50] Introducing Apache Mahout. [Online]. Available: https://www.ibm.com/developerworks/library/j-mahout/.

[51] Morales, G.D.F. and Bifet, A., 2015. SAMOA: scalable advanced massive online analysis. Journal of Machine Learning Research, 16(1), pp.149-153.

[52] Fahim, M. and Baker, T., 2017. Knowledge-Based Decision Support Systems for Personalized u-lifecare Big Data Services. In Current Trends on Knowledge-Based Systems (pp. 187-203). Springer International Publishing.

[53] Sodiya, A. S. Generic prediction of malaria treatmeant outcomes using big data analytics, 26th National Conference & Exhibition on Information Technology for National Safety & Security, pp. 75-83, 2016.

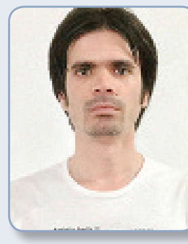
[54] Talia, D., 2015, July. Making knowledge discovery services scalable on clouds for big data mining. In Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2015 2nd IEEE International Conference on (pp. 1-4). IEEE.

[55] Chauhan, R., Kaur, H., Acharjya, D.P., Dehuri, S. and Sanyal, S., 2015. A spectrum of big data applications for data analytics. Comput Intell Big Data Anal, 19, pp.165-179.

약 력

무하마드 비랄 아민

2006년 미국 DePaul University 소프트웨어공학 석사
2015년 경희대학교 컴퓨터공학과 박사
2015년~2016년 경희대학교 박사후연구원
2016년~현재 경희대학교 연구교수
관심분야: 클라우드 기반 IoT, 성능 기반 어플리케이션, 분산시스템



무하마드 사딕

2012년 파키스탄 University of Peshawar 전산학 석사
2016년~현재 경희대학교 컴퓨터공학과 석사과정
2012년~2016년 파키스탄 Orbit Software Solutions 선임 소프트웨어 엔지니어
관심분야: 병렬 및 분산 온톨로지 매칭, 분산시스템, 지능형컴퓨팅

막불 알리

1999년 파키스탄 University of Agriculture 농업공학 학사
2002년 파키스탄 University of Agriculture 전산학 석사
2013년 파키스탄 National University of Sciences and Technology 정보기술학 석사
2014년~현재 경희대학교 컴퓨터공학과 박사과정
관심분야: 지식공학, 기계학습, 자연어처리, 사례기반학습

방 재 훈

2011년 평택대학교 디지털응용정보학과 이학사
2013년 경희대학교 컴퓨터공학과 공학석사
2013년~현재 경희대학교 컴퓨터공학과 박사과정
관심분야: 유비쿼터스 컴퓨팅, 음성처리, 기계학습, 감정인식