

Chapter 9

Knowledge-Based Decision Support Systems for Personalized *u*-lifecare Big Data Services

Muhammad Fahim and Thar Baker

Abstract The emergence of information and communications technology (ICT) and rise in living standards necessitate knowledge-based decision support systems that provide services anytime and anywhere with low cost. These services assist individuals for making right decisions regarding lifestyle choices (e.g., dietary choices, stretching after workout, transportation choices), which may have a significant impact on their future health implications that may lead to medical complications and end up with a chronic disease. In other words, the knowledge-based services help individuals to make a personal and conscious decision to perform behaviour that may increase or decrease the risk of injury or disease. The main aim of this chapter is to provide personalized ubiquitous lifecare (*u*-lifecare) services based on users' generated big data. We propose a platform to acquire knowledge from diverse data sources and briefly explain the potential underlying technology tools. We also present a case study to show the interaction among the platform components and personalized services to individuals.

Keywords Big Data Services · Personalized *u*-lifecare · Decision support system · Knowledge-based system

9.1 Introduction

We live in the age of data; everything surrounding us is linked to a data source (e.g., smartphone, smartwatch, wearable computing devices including smart glasses) that captures massive amount of valuable data digitally. This data can be utilized to

M. Fahim (✉)

Department of Computer Engineering, Faculty of Engineering and Natural Sciences,
Istanbul Sabahattin Zaim University, 34303 Istanbul, Turkey
e-mail: muhammad.fahim@izu.edu.tr

T. Baker

Department of Computer Science, Faculty of Engineering and Technology,
Liverpool John Moores University, Liverpool, UK

© Springer International Publishing AG 2017

G. Alor-Hernández and R. Valencia-García (eds.), *Current Trends on Knowledge-Based Systems*, Intelligent Systems Reference Library 120,
DOI 10.1007/978-3-319-51905-0_9

generate recommendations, monitor physical activities, and generate real-time alerts for different application domains such as sports, lifestyle, and healthcare. The same data is used with new datum to extract hidden trends, relationships, and association [1]. Processing data to extract knowledge is associated with many challenges due to the big data volume, speed through which it is generated, and redundancy and noise [2]. More importantly, we lack the time and capacity to study and search the collected data manually in real time in order to build knowledge that helps in steering our future. These challenges signify the need to the development of new technologies, tools, and practices for collecting, integrating, analyzing, and presenting a large volume of information to enable better decision making [3].

Consider a daily life routine scenario, where human interacts with multiple smart devices and requests personalized services for daily routines, softening the mood, healthy food or exercise routines. If we are able to process the data around human, and construct and manage the knowledge over the reliable IT infrastructure, then providing the personalized u -lifecare services at the right time becomes possible.

Single source of information is neither adequate nor sufficient to understand human, as human cognition comes from the knowledge of physical, mental, and social contexts. Furthermore, the behavior changes also have high impact to understand daily routines. Currently, the main challenge that remains under investigated is how to acquire knowledge from diverse data sources in efficient time and cost. In order to process such big data, acquire knowledge and build decision support system, we propose a platform that will accommodate diverse sources of structured and unstructured data and briefly explain the underlying technologies and tools to process the produced big data in a cost/time-effective way.

9.2 Related Work

Knowledge-based decision support systems play pivotal role in improving the quality of life by providing tools and services needed to resolve emergent problems, and knowledge necessary to suggest a new/specific strategy to work with various situations [4]. Many researchers and companies focused on exploiting the data for providing human-centric services, but their attempts are still limited. For instance, Nike+ [5], Samsung Gear [6], LG Smartwatch [7], Microsoft Band [8], Fitbit Blaze [9], to name but a few, promote active lifestyle and provide some basic health recommendations based on the recognized human activities, burned calories or hours of sleep. The same data can be used to extract hidden trends, relationships and association with wellness application, chronic disease prevention as well as analyzing the specific group of people.

Bilal et. al. [10] introduced data curation framework to accumulate users' sensory data from multimodal data sources in real time and preserve as a lifelog. They provide management support for the collection of large volume of sensory data. This data will be further processed under the Mining Minds (MM) platform [11] to generate knowledge and recommendation services for individuals.

Reza et al. [12] considered the smartphone as a portable computer. Such devices can be used for personal data collection because users carry their smartphone all the time. Authors also acknowledge the fact that digital data grow rapidly; therefore, it is difficult to process using traditional data management tools and techniques. Hence, Cloud Computing infrastructure along with big data processing provides more opportunities for critical infrastructure systems including health and human welfare, commerce and business, and economic systems, etc. with reduced cost [13, 14]. Miguel et al. [15] introduced a semantic platform for cloud services annotation and retrieval from their descriptions. The system can automatically annotate different cloud services from their natural language description.

The discovery of the knowledge typically involves the use of human expert knowledge and other source of information, which can be stored in logical structures, accessible and readable by machines [16]. Marco et al. [17] come up with semantic representation in terms of ontology-based knowledge to support some phases of the decision making process. The proposed approach has been successfully implemented and exploited in a decision support system for personalized environmental information. Similarly, another perspective of generating knowledge is introduced by Ling et al. [16]. They proposed knowledge generation by assisting the expert-driven rules via a hybrid method, which combines human domain expertise with machine learning methods to provide more accurate, effective and efficient method for discovering knowledge in complex domains.

In order to present the knowledge in a useful manner, there exist numerous challenges. Kambatla et al. [12] highlights those challenges in details to provide useful analytics. In terms of enhancing the wellbeing, data in healthcare poses some of the most challenging problems to large-scale integration as generated data is huge and contains verities including electronic medical records (EMR) and electronic health records (HER), imaging data, or personalized drug response. Similarly, several analysis tasks are also time-critical. For example, patient diagnoses, progression of outbreaks, etc., all have tight performance requirements. So, in this regard, a platform is required that can process such a big data and response in time. Our proposed platform keeps these challenges in mind and carefully takes care of all discussed issues in this section.

9.3 Proposed Platform

The proposed platform is based on big data storage and processing framework, data management, learning models, construction of knowledge bases, and personalized u -lifecare services. It is illustrated in Fig. 9.1, and details are given in the following sections.

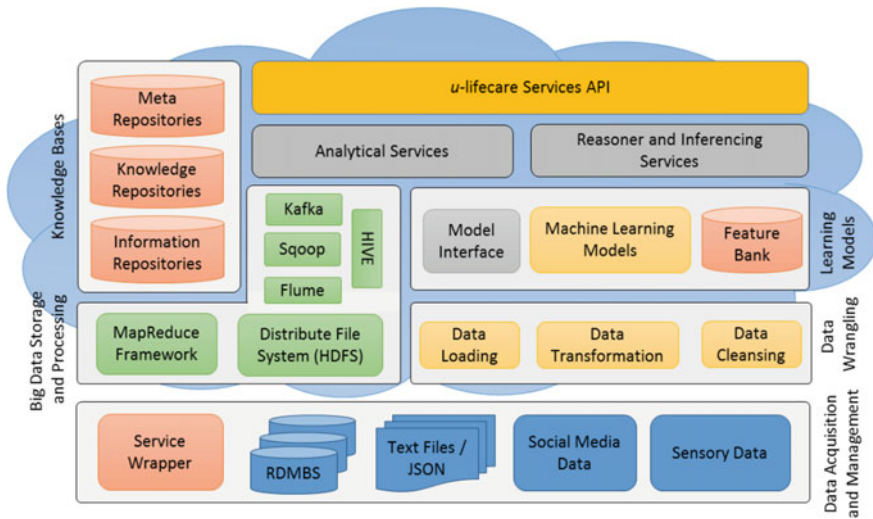


Fig. 9.1 The proposed knowledge-based decision support system for personalized *u-lifecare* big data services

9.3.1 Data Acquisition and Management

According to IBM, 90% of data in the world today has been created only in the last two years [18]. This data comes from heterogeneous data sources and have a lot of varieties including structured, unstructured and partially structured data, as shown in the bottom layer in Fig. 9.1. These data sources consist of multimodal physical sensors, which are based on embedded and different wearable devices such as smartphones and smart shirts [19–21]. It also includes the social networks data that is considered as an input source to our proposed platform. The input is raw data gathered and partially structured with respect to sensor or source categorization in a (csv, xml, dat, JSON, relational data, web scrapping, or text files, etc.). Each data source has its own configuration properties for establishing the connection and grabbing the existing or new data. We designed the data source “*service wrapper*” to collect and store all the technical information that is required to access the data from a particular source as well as data synchronization information. Once we are connected to a data source, we can gather the data and store the logs into Hadoop Distribute File System (HDFS) in real time.

9.3.2 Data Wrangling

One of the major issues encountered when building knowledge-based decision support systems is acquiring high quality of data and reshaping it for further process. In this regard, data wrangling technique shapes and manipulates the raw data to an

alternative format that is suitable for exploration and analysis [22]. In other words, it is able to convert and map data from raw form into a useful form that is more convenient for specific use. It includes cleansing, transformation and loading into target repository. Our consideration toward the data generation is big data that means generated data have high volume, velocity, variety and veracity. In this case, very well know technique so-called ETL (Extraction, Transformation and Loading) and models are not suitable because they require manual work from technical and domain experts at different stages of the process [22]. Data wrangling/munging is one the technique that works for both internal and external data sources with semi-automated tools and less human intervention. Details of data wrangling sub-modules namely (Data Cleansing, Data Transformation and Data Loading) are given below.

9.3.2.1 Data Cleansing

The gathered data from a single source, or multiple sources, is deemed correct but contains many inconsistencies and errors [23]. These issues arise because data is not collected for constructing the knowledge base and analysis purposes. In this case, we need to tackle the main issues (i.e., ambiguous or invalid data), missing values, duplicate entries, upper versus. lower case, date and time zones, etc. Treating these issues will increase the quality of data and assist in extracting the appropriate information and facts that lead to the effective decision-making. In order to get the unified format of data, we need to prepare the data in a consistent way and solve the above-mentioned main issues associated with the collected data. In Table 9.1, we

Table 9.1 Solutions to the issues along with illustrative examples

Issues	Examples	Unified format
Inconsistent data	Gender can be M/F or 1/0 or male/female	We need to define the rule, which one should we keep for the further processing of data
Ambiguous or invalid data	Last updated: 2026	This data is considered as invalid, and we can find outliers by visualizing the histogram
Missing values	Sensor reading: _____	We can estimate the missing attribute value based on other examples for which this attribute has a known value Simple way to assign the most common value for that attribute or delete the instance depending on the situation. A more complex technique is to assign a probability to each of the possible values of attributes [24]
Duplicate Entries	User ID field in sensor logs and social media	If we merge these data files then we can keep the single User ID instead of two
Upper versus. lower case	Sitting or sitting	We can define rules for handling case-sensitive matters as well
Date format and time zones	Date: 7/29/2016 13:15 Date: 29/7/2016 12:15 pm	We can define rule for consistent date format and time zone

provide the possible solutions to the issues along with examples for the transformation of data into a unified format.

We can use some existing tools for data cleansing to ease our job such as Data Wrangler [25], Tabula [26], OpenRefine [27], Python and Pandas [28], to name a few.

9.3.2.2 Data Transformation

In data transformation, we will apply the smoothing technique in order to remove the noise and short-term irregularities hence improve accuracy of forecasts and readings. Smoothing also have more positive effects on important processes in data transformation such as pre-joins given that joins are expensive and warrant special performance consideration as they may create data sets beyond what it is needed for so pre-joining data once and store them for further future use deemed requisite; normalization to scale the data in the specified range, discretization by dividing the range of continuous attribute into intervals, and new attribute construction from the given one. The following table highlights the data transformation requisites and corresponding formulas and explanation (Table 9.2).

9.3.2.3 Data Loading

After applying cleansing and transformation techniques, we need to load this structured and partially structured data for the efficient access whenever requested by the other components to process this data. The temporal data is stored in HDFS by utilizing a distributed “Apache Flume” data service [29]. Flume consists of agents, which includes a source, channel, and sink; which all work together to flow data from data sources to the required destination. HDFS has many advantages over the others as open source, scalable, reliable, manageable and moving large amounts of log data, amongst others.

Table 9.2 Transformation formulas

Transformation	Formula/Method	Explanation
Normalization	$Z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$	Where $x = (x_1, \dots, x_n)$ and Z_i is i th normalized data between the range of 0 and 1
Discretization [26]	Threshold method Multi-interval method Gain Ratio method	Discretizing the continuous values attribute into multiple intervals
New attribute construction	Mean, standard deviation, Variance	Statistically attributes help to know more about the data

9.3.3 *Big Data Storage and Processing*

Data comes from the heterogeneous sources including sensors (i.e., embedded in smartphone, smartwatch or wearable devices), social networks, publically available datasets, and historical data for extracting the valuable information. Consequently, the size, velocity, variety and veracity of the data, is huge and hence requires reasonable time and cost for the processing. Thus, our proposed data storage is based on an open source framework Apache Hadoop that supports data intensive jobs [30]. Hadoop framework has a master–slave architecture and built-in fault tolerance capability by making three or user defined replica of data nodes. Our platform provides the high performance computing over the commodity hardware of cloud infrastructure OpenStack [31]. Our platform utilized MapReduce framework that is a distributed, parallel processing architecture and uniquely qualified to exploit big data potential [32]. A MapReduce job comprises of two parts, (i) a map part, which takes raw data and organizes it into key/value pairs, and (ii) a reduce part, which processes data in parallel. This component also contains tools for data reading, writing, movement and interaction, such as Flume [29], kafka [33], Sqoop [34] and Hive [35].

9.3.4 *Learning Models*

The proposed learning models component is capable to learn from the large collected data and use that knowledge to predict future trends, behaviors, and decisions regarding unknown future events. Learning models contains feature bank and machine learning models. In our feature bank component, we extract the relevant feature according to the data source. Feature extraction is a highly domain-specific technique that defines a new attribute using the raw signals to reduce computational complexity and enhance the recognition process. For instance, in case of accelerometer signal we extracted the time and frequency domain features [36]. Learning model contains standard algorithms of recognizing the human contexts and behavior including non-parametric nearest neighbor model [37], evolutionary fuzzy model [38], social media processing API [39] for classification and prediction. The details about the learning models can be found in our recent publications [40].

At this point, data will be transforming to meaningful information and can provide directly to analytical service for visualizing the human contexts and behavior patterns. In our big data processing approach, machine-learning models perform batch processing in which all training data set is read once and learned parameters are stored in knowledge repositories. We stored parameters in knowledge bases instead of HDFS because frequent I/O operations can become very expensive in terms of time.

9.3.5 Model Interface

We are introducing the model interface that represents a standard interface to provide the linkage between components. It can communicate with knowledge bases as well as communication between big data storage and processing component for getting the training data. When new data arrive to the platform, or users want to retrain the learning model parameters, it will be coordinated with model interface. We can move learning models information and knowledge repositories into HDFS if the size increases from gigabytes to petabytes. This component is also responsible to maintain and modify the learning parameters when new data dimensions arrive.

9.3.6 Knowledge Bases

Once the data is distilled and processed through learning models it is loaded into information repositories, so users have cost effective and real time access to it. It also contains the information, knowledge, and Meta repositories. Knowledge repositories are filled after the inferencing module of our platform. Knowledge repositories assist the analytical services for visualization as well as reasoner and inferencing services for better quality of decisions as well as reasoning about the certain situations. In case of Meta repositories, it contains the metadata that contains the schema information and information about the integrity constrains. It will be helpful for the other APIs' consumer engineers.

9.3.7 Reasoner and Inferencing Services

This component assists the platform to provide personalized recommendations and reasoning for provided guidelines. Traditionally, reason and inference module focus on general recommendations applicable to a community of users, but not specific to each individual and their personal preferences. To accommodate personalization concept and dynamic user's query support at run time, a hybrid reasoning architecture is proposed, exploiting different approaches, such as rule-based reasoning [41], preference-based reasoning [42] and probabilistic inferencing [43]. Initially, rules are extracted explicitly from the domain knowledge and coarse grain by user preferences. To make final decision probabilistic, Bayesian network is utilized to gain certain confidence. The system can be tuned by the preference of the user collected at the initial configuration time that whether to use a single or a combination of the approaches or a specific reasoning method to generate the prompt response to their queries.

9.3.8 *Analytical Services*

Analytical services provide the visualization and new insight of data to uncover hidden patterns, unknown correlations and users' behavior. In certain situations, experts need current information while in other scenarios they desire historical information along with the current information. We placed Hive-based queries looking inside the data and analytical services by providing the web interfaces. The tools include d3 [44], ggplot [45], matplotlib [46], and Google charting [47], to mention just some. It can support the experts to prepare the better and effective recommendation plan for the users.

9.3.9 *u-Lifecare Services API*

The decision support system can be accessed via a user interface and service API to build the requested applications over the constructed knowledge base. The objective of personalized *u*-lifecare services is to provide timely and accurate services to the individuals based on the constructed knowledge, user's generated data as well as historical data. This represents the top layer in Fig. 9.1, and linked directly to the analytical services, and reasoner and inferencing services.

9.4 Case Study

Consider the monitoring and tracking of user's behaviour routine of a focused group or nationwide. User's behaviour can be divided into active or sedentary. While sedentary behaviour is increasing due to societal changes and related to prolonged periods of sitting. Sitting while watching television, using the computer while working or playing games for long hours, are examples of sedentary behaviours that are currently common worldwide [48]. These kinds of activities increase sedentary behaviour across all age groups. A person is considered sedentary if they spend large amount of their day with such activities and do not spend sufficient time for physical activity or exercise. Similarly, many jobs require people to sit in front of computer all the day, which also promotes sedentary behaviour. Sedentary behaviour is associated with poor health outcomes, including the high risk of overweight and obesity [49], physiological and psychological problems [50], heart disease and diabetes [51]. To promote healthy behaviour, there should be some efficient mechanisms to track and estimate the time spent in active and sedentary activities. In order to track user's behaviour in daily routines, we developed fundamental contexts tracking application based on embedded sensor of smartphone [38, 52]. Our application is capable to run in the background while users can use their smartphone for other tasks. We construct the feature bank by extracting the

relevant features according to the type of sensor. During the classification phase, stored features from the codebook are loaded into the learning model and classify the current situation. Figure 9.2 shows example scenes of fundamental contexts of human behaviour.

For discussion, consider the weight management scenario for people, like Ms. Aliza. She is a 28-years-old lady, who wants to adopt an active lifestyle in her daily routines. She preferred physical exercise, such as brisk walking and jogging. Her recent weight gain has prompted her to adopt physical activities in daily routines, which can be squeezed into daily schedule with ease. She needs guidelines and recommendations that fit in her busy schedule. Statistics about her body mass index (BMI) is summarized in Table 9.3.

Aliza is interested to know about her lifestyle primitive statistics, which tells her how much physical activity or sedentary behaviour she did in the previous days as well as recommendations and routine plan in terms of daily routines to achieve her active lifestyle goal. Supporting this scenario, the proposed platform can help her in a truly ubiquitous manner to log her daily routines. She installed our developed application and creates her profile. She can maintain her personal profile; change the preferences according to the seasonal changes and system can generate personalized recommendation according to new preferences. Our application can run in the background and log the routines while using the smartphone normally. She can share her workouts over the social networks with family and friends. The sharing feature of our platform is optional and has great potential to motivate Aliza in terms of appreciation from the social networks while promoting an active



Fig. 9.2 Example of active behavior ‘Stairs’, ‘Running’, ‘Jogging’, ‘Cycling’, ‘Walking’ and example of sedentary behavior ‘short break’, ‘working on PC’, ‘watching TV’, and ‘Sedentary—Unknown Activity’

Table 9.3 User’s physical statistics

Gender	Age (Years)	Height (cm)	Weight (lbs)
Female	28	162	160

Table 9.4 Twenty four hour routine with time spent duration

Activity	Duration (h)
Active	3.75
Walking	0.30
Stairs	0.05
Cycling	0.15
Short breaks	0.50
Working on PC	3.58
Watching TV	1.55
Sedentary—context unknown	14.12

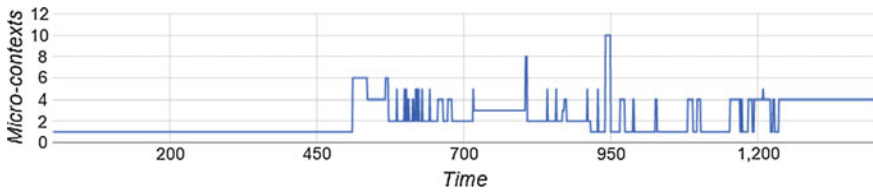


Fig. 9.3 Daily routine with time spent of activity pattern

lifestyle. Consequently, it motivates the other individuals to adopt healthy routines to keep themselves healthier. We are presenting her 24 h data of a working day. Table 9.4 shows one-day activities along with micro-contexts.

In Table 9.4, we showed the quantification of the amount of time spent in sedentary behavior by subject that was around 19 h and 15 min. In the contexts class label “Sedentary—context unknown”, we consider the subject sleeping time and all other micro-contexts such as subject went to library for studying or any other that is not included in the recognized micro-context. Even though short break time is 30 min but important to indicate the state of being active. In order to get insight of the sedentary patterns of daily routines, we showed observed pattern from 12:00 to 23:59 by presenting each minute of 24 h in Fig. 9.3, where x-axis shows the time in minute while y-axis shows the micro-context and annotation as shown in Table 9.5.

Table 9.5 Twenty four hour routine with time spent duration

Activity	Class label
Sedentary	1
Working on PC	2
Watching television	3
Active	4
Short break	5
Walking	6
Running	7
Stairs	8
Jogging	9
Cycling	10

This micro-context of sedentary behaviour provide better understanding of users' daily routines and may help users to minimize the amount of prolonged sitting and adopt active lifestyle. We are using Google charting [34] for presenting behaviour analytics.

In Fig. 9.3, the subject activity is sedentary because of night and subject was sleeping. After that we can see active patterns, using computer, watching television and on off state of active and sedentary. We classify "short breaks" if subject's activity time is less than or equal to one minute and "active" if time is more than one minute. For instance, we also show the activities along starting time and sequence in Fig. 9.4.

In Figs. 9.5 and 9.6, daily routine pattern is visualized and provide us information on percentage of time in 24 h spent in different micro-contexts activities. We also extract the information about the number of short breaks (i.e., 15 short breaks) that help subject to avoid longer sedentary activity.

Our platform also provides recommendations with the help of reasoner and inference service module. It correlates her personal profile along preferences with daily routines and suggest routine plan to adopt healthy behaviour. For instance, if reasoner assesses Ms. Aliza in prolonged sitting state while watching television. Our platform generates the appropriate physical activity to complete the routine plan of the users. The inference service recommends her the following recommendations over the smartphone as a toast message with beep.

To be active, you can take a brisk walk of 15 min. Brisk walking helps you to reduce body fat, depression, and give you stronger bones.

She can also query about the weather conditions or any further recommendations that she require to go out for recommended activity. The execution process of our platform is illustrated in the sequence diagrams of Fig. 9.7.

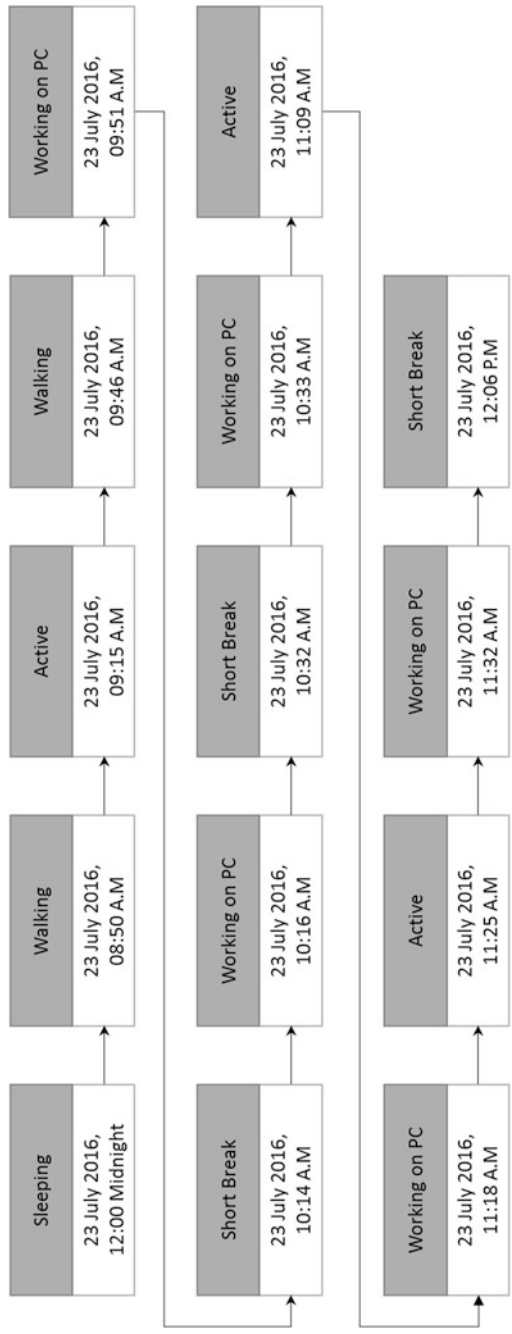


Fig. 9.4 Context recognition along with activity starting time and sequence of occurrences for 12 h

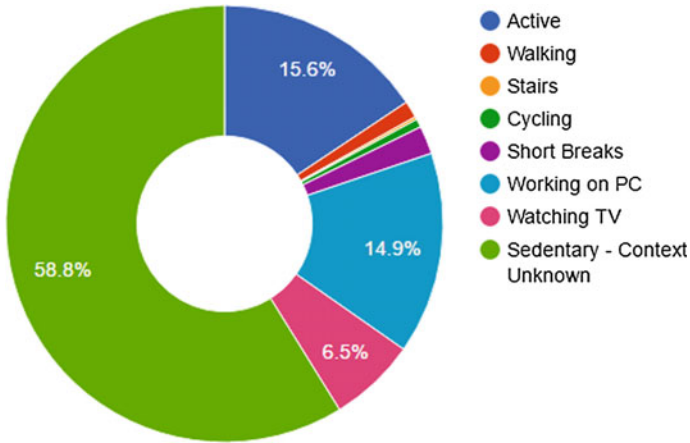
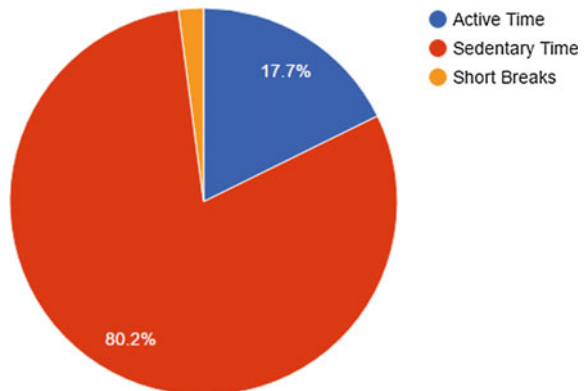


Fig. 9.5 Sedentary behaviour along percentage of time spent in different activities

Fig. 9.6 Sedentary behaviour with three broad categories of 24-hour routine



Our behaviour recognition analysis may help the users to minimise the amount of time spent in prolonged sitting and encourage them to break up long periods of sitting as often as possible. We can quantify the time spent in electronic media during leisure time (e.g., television, video games and computer use) and set the limit hour of their usage. We can also include the social networks to know the activities and recommend the user’s active group on social media, to get more knowledge, and educate own selves while overcoming barriers of physical distance or geographic isolation. It will consequently help to adopt the healthier lifestyle and reduced the health risks.

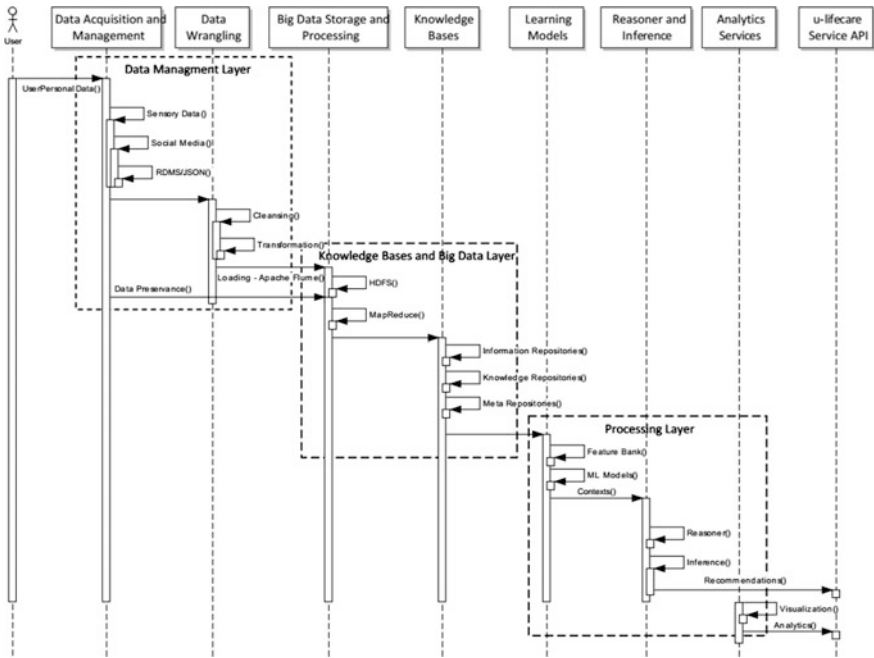


Fig. 9.7 Sequence diagram of execution process in case study

9.5 Conclusion

We proposed a platform that is able to process the structured and unstructured data gathered from multiple sources through big data technology, to provide consolidated services and analytics to assist in decision making. Our proposed platform includes number of modules and sub-modules that are capable of data collection and management, extract the knowledge and information while exploiting big storage technology for massive storage of data. For continuous sensing and acquiring of personal data, we utilized ubiquitous nature of cloud computing infrastructure. Furthermore, it provides high performance computing for intensive data processing in cost effective manner. The proposed platform proved systematic data management and effective utilization of the users' generated data that can help the individuals to visualize the personal behaviour patterns and provided u-lifecare services to manage their daily routines and remain active. Our future plan includes providing a comparative analysis for individuals based on certain parameters while preserving the privacy and security of the data.

References

1. Brule, M.R.: Big data in exploration and production: real-time adaptive analytics and data-flow architecture. In: SPE Digital Energy Conference. Society of Petroleum Engineers (2013)
2. Provost, F., Fawcett, T.: Data science and its relationship to big data and data-driven decision making. *Big Data* **1**(1), 51–59 (2013)
3. Dayal, U., Castellanos, M., Simitsis, A., Wilkinson, K.: Data integration flows for business intelligence. In: Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, pp. 1–11. ACM (2009)
4. Shen, J., Xing, J., Xu, M.: Research on CBR-RBR fusion reasoning model and its application in medical treatment. In: Proceedings of the 21st International Conference on Industrial Engineering and Engineering Management, pp. 431–434 (2015)
5. [Nike+]. https://www.nike.com/us/en_us/c/nike-plus/nike-app (2016). Accessed 30 Sept 2016
6. [Samsung Gear]. <http://www.samsung.com/us/mobile/wearables/> (2016). Accessed 30 Sept 2016
7. [LG Smartwatch]. <http://www.lg.com/us/smart-watches> (2016). Accessed 30 Sept 2016
8. [Microsoft Band]. <https://www.microsoft.com/microsoft-band/en-us> (2016). Accessed 30 Sept 2016
9. [Fit Blaze]. <https://www.fitbit.com/blaze> (2016). Accessed 30 Sept 2016
10. Amin, M.B., Banos, O., Khan, W.A., Muhammad Bilal, H.S., Gong, J., Bui, D.M., Cho, S.H., Hussain, S., Ali, T., Akhtar, U., Chung, T.C.: On curating multimodal sensory data for health and wellness platforms. *Sensors* **16**(7) p. 980 (2016)
11. Banos, O., Amin, M.B., Khan, W.A., Afzal, M., Hussain, M., Kang, B.H., Lee, S.: The mining minds digital health and wellness framework. *BioMedical Engineering OnLine* (2016)
12. Rawassizadeh, R., Tomitsch, M., Wac, K., Tjoa, A.M.: UbiqLog: a generic mobile phone-based life-log framework. *Pers. Ubiquit. Comput.* **17**(4), 621–637 (2013)
13. Kambatla, K., Kollias, G., Kumar, V., Grama, A.: Trends in big data analytics. *J. Parallel Distrib. Comput.* **74**(7), 2561–2573 (2014)
14. Rodríguez-Mazahua, L., Rodríguez-Enríquez, C. A., Sánchez-Cervantes, J. L., Cervantes, J., García-Alcaraz, J. L., Alor-Hernández, G. A.: General perspective of Big Data: applications, tools, challenges and trends. *J. Supercomputing* pp. 1–41 (2015)
15. Rodríguez-García, M.Á., Valencia-García, R., García-Sánchez, F., Samper-Zapater, J.J.: Creating a semantically-enhanced cloud services environment through ontology evolution. *Future Gener. Comput. Syst.* **32**, 295–306 (2014)
16. Ling, T., Kang, B. H., Johns, D. P., Walls, J., Bindoff, I.: Expert-driven knowledge discovery. In: IEEE Fifth International Conference on Information Technology: New Generations pp. 174–178 (2008)
17. Rospocher, M., Serafini, L.: An ontological framework for decision support. In: Joint International Semantic Technology Conference pp. 239–254 (2012)
18. [IBM]. <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html> (2016). Accessed 8 July 2016
19. Gupta, P., Dallas, T.: Feature selection and activity recognition system using a single triaxial accelerometer. *IEEE Trans. Biomed. Eng.* **61**(6), 1780–1786 (2014)
20. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. *ACM SIGKDD Explor. Newslett.* **11**(1), 10–18 (2009)
21. Theekakul, P., Thiemjarus, S., Nantajeewarawat, E., Supnithi, T., Hirota, K.: A rule-based approach to activity recognition. In: Knowledge, Information, and Creativity Support Systems, pp. 204–215. Springer, Berlin (2011)
22. Furche, T., Gottlob, G., Libkin, L., Orsi, G., Paton, N.W.: Data wrangling for big data: challenges and opportunities. In: 19th International Conference on Extending Database Technology (EDBT). Bordeaux, France (2016)
23. Rahm, E., Do, H.H.: Data cleaning: problems and current approaches. *IEEE Data Eng. Bull.* **23**(4), 3–13 (2000)

24. Michalski, R.S., Carbonell, J.G., Mitchell, T.M. (eds.): *Machine Learning: An Artificial Intelligence Approach*. Springer Science & Business Media (2013)
25. [Data Wrangler]. <http://vis.stanford.edu/wrangler/> (2016). Accessed 28 July 2016
26. [Tabula]. <http://tabula.technology/> (2016). Accessed 28 July 2016
27. [Openrefine]. <http://openrefine.org/> (2016). Accessed 28 July 2016
28. [Python and Pandas]. <http://pandas.pydata.org/> (2016). Accessed 28 July 2016
29. [Apache Flume]. <http://flume.apache.org/> (2016). Accessed 28 July 2016
30. [Apache Hadoop]. <http://hadoop.apache.org> (2016). Accessed 28 July 2016
31. Sefraoui, O., Aissaoui, M., Eleuldj, M.: OpenStack: toward an open-source solution for cloud computing. *Int. J. Comput. Appl.* **55**(3) (2012)
32. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. *Commun. ACM* **51**(1), 107–113 (2008)
33. [Apache Kafka]. <http://kafka.apache.org/> (2016). Accessed 28 July 2016
34. [Apache Sqoop]. <http://sqoop.apache.org/> (2016). Accessed 28 July 2016
35. [Apache Hive]. <http://hive.apache.org/> (2016). Accessed 28 July 2016
36. Fahim, M., Lee, S., Yoon, Y.: SUPAR: Smartphone as a ubiquitous physical activity recognizer for u-healthcare services. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3666–3669 (2014)
37. Fahim, M., Khattak, A.M., Chow, F., Shah, B.: Tracking the sedentary lifestyle using smartphone: a pilot study. In: 2016 18th International Conference on Advanced Communication Technology (ICACT), pp. 296–299 (2016)
38. Fahim, M., Fatima, I., Lee, S., Park, Y.T.: EFM: evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer. *Appl. Intell.* **39**(3), 475–488 (2013)
39. [Alchemy API]. <http://www.alchemyapi.com/social-media-monitoring> (2016). Accessed 28 July 2016
40. Fahim, M., Idris, M., Ali, R., Nugent, C., Kang, B., Huh, E.N., Lee, S.: ATHENA: a personalized platform to promote an active lifestyle and wellbeing based on physical, mental and social health primitives. *Sensors* **14**(5), 9313–9329 (2014)
41. Sun, R.: Robust reasoning: integrating rule-based and similarity-based reasoning. *Artif. Intell.* **75**(2), 241–295 (1995)
42. Ali, R., Afzal, M., Hussain, M., Ali, M., Siddiqi, M.H., Lee, S., Kang, B.H.: Multimodal hybrid reasoning methodology for personalized wellbeing services. *Comput. Biol. Med.* **69**, 10–28 (2016)
43. Murphy, K.P.: *Dynamic bayesian networks: Representation, inference and learning* (Doctoral dissertation, University of California, Berkeley) (2002)
44. [D3]. <https://d3js.org/> (2016). Accessed 28 July 2016
45. [ggplot2]. <http://ggplot2.org/> (2016). Accessed 28 July 2016
46. [matplotlib]. <http://matplotlib.org/> (2016). Accessed 28 July 2016
47. [Google charting]. <https://developers.google.com/chart/> (2016). Accessed 28 July 2016
48. Dunstan, D.W., Healy, G.N., Sugiyama, T., Owen, N.: Too much sitting: the population health science of sedentary behavior. *Eur. Endocrinol.* **6**(1), 19–23 (2010)
49. Tremblay, M.S., Colley, R.C., Saunders, T.J., Healy, G.N., Owen, N.: Physiological and health implications of a sedentary lifestyle. *Appl. Physiol. Nutr. Metab.* **35**(6), 725–740 (2010)
50. Barwais, F.A., Cuddihy, T.F.: Empowering sedentary adults to reduce sedentary behavior and increase physical activity levels and energy expenditure: a pilot study. *Int. J. Environ. Res. Public Health* **12**(1), 414–427 (2015)
51. Vandelandotte, C., Duncan, M.J., Short, C., Rockloff, M., Ronan, K., Happell, B., Di Milia, L.: Associations between occupational indicators and total, work-based and leisure-time sitting: a cross-sectional study. *BMC Public Health* **13**(1), 1 (2013)
52. Fahim, M., Khattak, A.M., Thar, B., Chow, F., Shah, B.: Micro-context recognition of sedentary behaviour using smartphone. In: 2016 6th International Conference on International Conference on Digital Information & Communication Technology & its Applications (DICTAP2016) (2016)