A wearable device-based personalized big data analysis model

Shujaat Hussain[1], Byeong Ho Kang[2] and Sungyoung Lee[1]

[1]Department of Computer Engineering, Kyung Hee University, Korea,
{shujaat.hussain, sylee} @oslab.khu.ac.kr
[2]Department of Science, Engineering and Technology University of Tasmania,
Australia ABN 30 764 374 782,
byeong.kang@utas.edu.au

Abstract. Wearable devices and the data generated by them gives a unique opportunity to understand the user behavior and predict future needs due to its personal nature. In coming years this data will grow exponentially due to huge popularity of wearable devices. Analysis will become a challenge with the personal data explosion and also to maintain a updated knowledge base. This calls for big data analysis model for wearable devices. We propose a big data analysis model which will update the knowledge base and give users a personalized recommendations based on the analysis of the data. We have designed a personalized adaptive analysis technique for data handling and transformation. This technique also responds to information utilization APIs in a real time manner. We are using mapreduce as our big data technology and ensure that data can be used for long term analysis for different applications in the future.

1 Introduction

In todays world a lot of data is generated by the wearable devices which is not leveraged and processed properly. This data is highly unique due to the personal nature of the wearable device. There is an exponential growth in wearable devices and they collect a lot of data which cannot be processed using conventional techniques. The target data are all the devices revolving around the user which can vary from the smartphones to glasses to wearable clothing. The data gathered from the wearable devices have yet to be properly explored or analyzed to enable the user to make decisions relevant to it.

Nearly 15 million wearable smart devices will be sold this year, amounting to $800 million [1]. All the major sensory data analysis is performed at remote systems, because the capability of a wearable or sensory device is premature to perform analysis on the periodic data. Data storage, their computation limit and multi-tasking compels to use the data at remote servers and then keep the data there at large scale. The same data is used then with new datum to extract hidden trends and states from both historical and their association. There is a lack of techniques and approaches which exploit the sensory data collected by wearable devices and use them for decision making and recommendations.

Also with the constant monitoring by the devices, there is a need for a big data solution which has the ability to analyze the huge amount of data.

Personalized Big data is becoming a reality with wearable computing on the rise. The wearable devices cover but are not limited to medical devices, sensor based wrist wear, headset, glasses, smartphone and clothing. All these devices collect a lot of data through user input and sensors. The next frontier in generating and leveraging data will be wearable computing. They generate data in real time which brings the big data into picture as it is equipped to handle these things. Big data refers to data that grows large in terms of velocity, volume, variety and veracity which mean that conventional techniques and database management tools are not as efficient or effective. The challenge arises from data volume, the speed through which is generated and redundant/noisy data. Big data techniques comprise of new tools which tackle the data deluge through parallelism, redundancy, machine learning and pattern recognition techniques and accommodate massive data analysis through big data infrastructure.

The recent developed devices related to the Human interaction like Googles Glass [2], Pebble [3], Fitbit One [4], heart beat monitor, smartphone, smart shirt etc, show the importance and opens new ways of research and application for IT and media development. Different companies are coming to the market at different angles like wearable as smartphones, pairing with Bluetooth, and applications.

Wearable technologies use sensors that generate sensory data which can be automatically detected, collected, categorized and analyzed. Sensory data can be associated with new datum to learn from the environment. In health care, sensory data from wearable as well as smartphone sensors is used to generate recommendations, monitor physical health activities, generate real-time alerts, and motivate the users for a quick and on time preparation[5]. Sensors are also used at large scale in buildings, institutions, and bridges to monitor their conditions humidity, temperature, heat dissipation and maintenance problems. Sensory data is collected and then based on historical analysis and learning models, alerts are generated to cater pre-cautions, before any damage or collapse [6]. In this paper a system is developed which acquires the data collected by the wearable devices and exploit them according to the user requirements and needs. An adaptive analysis model is created where user can define his requirements and data will be extracted according to those demands and presented in a way to enable the user to make decisions. The data will be stored in Big Data repository which will accommodate the data deluge generated by wearable devices. This will ensure that through this approach, analysis of huge dataset is done. We are using hadoop mapreduce [7][8] as our big data technology. The remainder of the paper is organized as follows. Section 2 briefly reviews related work, section 3 proposes a framework for a big data analysis model for wearable devices. Section 4 describes the experimental setup and section 5 is conclusion and future work.

2 Related Work

One of key differences between wearable devices and conventional devices is that the human and the devices are intertwined with each other which give a huge advantage in contextual situations. Wearable computing is leading towards personalized Big Data and is a basis for new platform. All the sensors in either of the forms are collecting lot of data, low latency analytics and enabling data visualization. The foundation of personalized big data is predictive search and it is entering a new stream. Major big companies like Apple, Google and Samsung are working on predictive search applications [9]. Even the startups and medium level companies are starting to show huge interest as the consumers are highly interested in these gadgets.

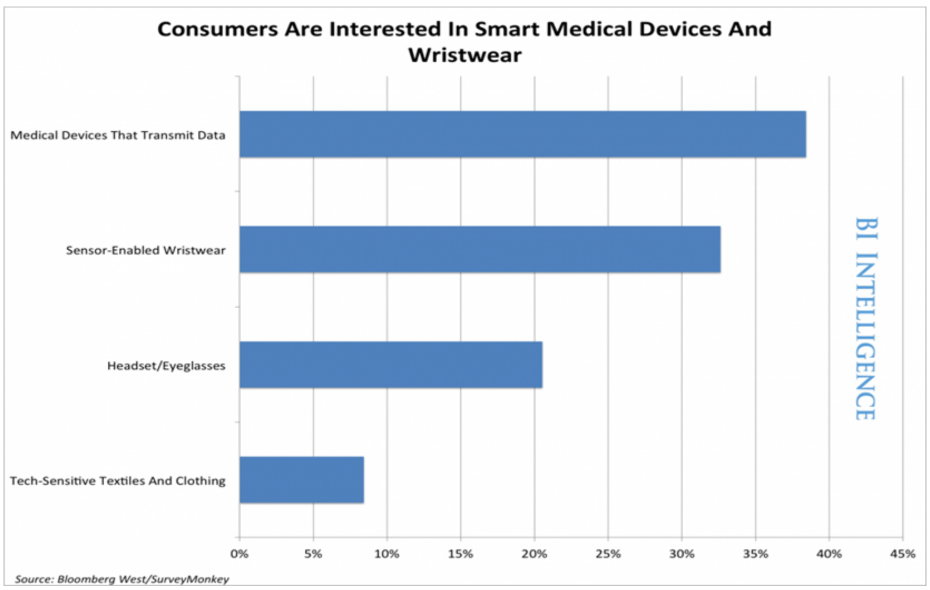

Fig. 1. Key Wearable Devices

Figure 1 [10] above shows the consumer interest in different kind of wearable devices. Medical wearable devices can play a big part in health monitoring and improvement. The most popular wearable device is the wrist wear and the headset/eyeglasses. It has been mentioned in [11] that huge costs is incurred due to underutilized data , so there is a strong need to exploit the data for user personalization. A more data driven approach is needed to properly use the data generated by the wearable devices. It has been emphasized in [12] that knowledge driven approach is also needed for real time continuous activity recognition in a multisensory environment. Data driven decision making can be effectively achieved by big data technology. Data driven decision making is referred as the practice of making decision based on the analysis of data rather than educated guess work and intuition[13]. With this analysis of data one other thing which is of paramount importance is the user's personal approach and providing him with right kind of information. For this there must be customized user based configurations[14].

3 Proposed Framework

Our focus is around the maintenance of the personalized knowledge base of the users and give recommendation based on the data transferred to the big data repository. Usually the main focus is to create a tool that generates personalized knowledge base but today it changes frequently due to different data gathering sources. We extract knowledge from big data and using expert system techniques to create and inquire heuristic knowledge of user and maintain/update this knowledge regularly and incrementally. The proposed system has three main components i.e. the data acquisition and data management, personalized adaptive analysis and personalized adaptive service layer as shown in figure 2. The details are below

3.1 Data Acquisition and Management

Data Acquisition The data acquisition passes the raw data to data transformation component. The data will be gathered through the android app and then connected to the cloud. The data acquisition takes input from the physical sensors which are attached to different wearable devices like accelerometer, GPS, light etc. The data is passed in an archived from the android application and sent to the data acquisition component.

Data Transformation The input is the raw data gathered in the data acquisition component. The data transformation layer takes the raw data and partially structure with respect to sensor categorization in a csv or text files. The streaming data is archived so that fast and light communication can be done between the cloud and the wearable device app.

Data validation The input is the partial structured data which is passed by the transformation component. The output is to crosscheck the sensor data and remove the redundancies from it. It is then stored in the Hadoop distributed file system. Data validation is done through finding holes in the continuous data timestamps and where the data is not consistent. In this way data will be aligned for training.

3.2 Personalized Adaptive Analysis

The adaptive analysis model will be adaptive and dynamic depending on the user needs and configuration.

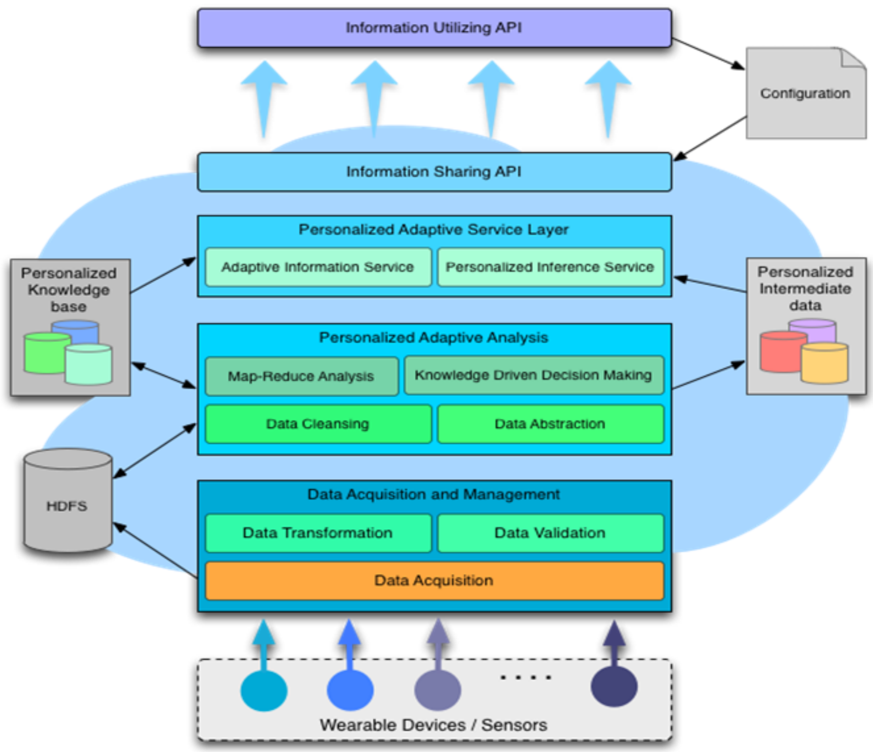

Fig. 2. Proposed Architecture

Data Abstraction The data abstraction selects user attributes from the configuration file which is processed further. The input of the data is partially structured data that is stored in the HDFS. The data abstraction selects user attributes from the configuration file which is be processed further. The configuration file is populated by the user through the information utilizing API

Data Cleansing The data cleansing module check the user attributes, their values and structure the schema to store in the personalized intermediate database. The user attributes are passed from data abstraction. The personalized intermediate database is being used to give a real time response to the information utilizing APIs. Querying the HDFS and bringing the relevant data takes a lot of time so an intermediate database is introduced.

Mapreduce Analysis Hadoop is a cloud computing platform and an open source implementation of MapReduce programming model [8]. In a MapReduce job there are three phases i.e. map, copy and reduce. The structured schema is passed on the mapreduce analysis component as well as the raw logs stored in HDFS. The output is be the historical values of the user attributes and produce training data for the decision making. The mapreduce analysis is an intermediate step to populate the personalized knowledge base. The SQL like queries are used for retrieving the data from the HDFS through Apache PIG[15].

Knowledge driven Decision making This module populate personalized knowledge base through trained parameters which will assist user in decision making. The MapReduce analysis with give the training data to knowledge driven decision making. The output is personalized knowledge base through trained parameters. We are using two machine learning techniques for this framework i.e. nave Bayesian and decision trees.

3.3 Personalized Adaptive Service Layer

Adaptive Information Service The adaptive information service will use intermediate data to expose user infographics to the user. This will assist user in monitoring his activities in an efficient manner. This information is structured from the raw data acquired and shows statistics in different visualized cues.

Personalized Inference Service Personalized inference service will use personalized knowledge base to get the facts populated by the knowledge driven decision making and give recommendations to the user. These recommendations will vary from user to user based on the activities and the wearable device.

4 Experimental Setup

We are using two server machines each with 16 GB ram and 8 cores. We installed four virtual machines (VMs) on server 1 and 2 VMs on server two. We are using an android application in the wearable device which archives the data and pass the data through a web service. This data is stored in Hadoop distributed file system. We access the data through Apache Pig scripts. The mapreduce programs are written in java and run on 6 VM hadoop cluster. We populate the intermediate database through these scripts and the adaptive knowledge base through mapreduce programs. There are two types of mapreduce implementation. One is for the analysis and one is the machine learning implementation in which decision making module is trained and tested.

5 Conclusions

Wearable data provides a unique opportunity because the user and the devices have a more intimate connection due to which we can get more relevant and contextual information. The data in the next few years is going to balloon and it is going to be a challenge to analyze and keep an updated knowledge base. Personalized Big Data is the next wave and associating different data sources will be key to personalization. All of the devices are going to collect a lot of data and present a lot of data in real-time which is an interesting use case for analytics. One of the Key technologies is mapreduce and its open source solution Hadoop. We have proposed a framework to give a real time solution to this problem and at the same time update the knowledge base in a smooth and efficient manner.

6 Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2014-(H0301-13-2001)) This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2013-067321).

References

1. wearablesdevcon: http://www.wearablesdevcon.com/article1.aspx (2014)
2. Google: www.google.com/glass/start/ (2014)
3. Pebble: http://en.wikipedia.org/wiki/pebble$_{(}watch)$(2014)
4. Fitbit: http://www.fitbit.com/kr/one (2014)
5. Pantelopoulos, A., Bourbakis, N.G.: A survey on wearable sensor-based systems for health monitoring and prognosis. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on **40** (2010) 1–12
6. Lorincz, K., Malan, D.J., Fulford-Jones, T.R., Nawoj, A., Clavel, A., Shnayder, V., Mainland, G., Welsh, M., Moulton, S.: Sensor networks for emergency response: challenges and opportunities. Pervasive Computing, IEEE **3** (2004) 16–23
7. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. Communications of the ACM **51** (2008) 107–113
8. Bhandarkar, M.: Mapreduce programming with apache hadoop. In: Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on, IEEE (2010) 1–1
9. Marrs, M.: http://www.wordstream.com/blog/ws/2013/06/24/predictive-search (2013)
10. Kalakota, R.: http://practicalanalytics.wordpress.com/2013/08/12/predictive-search-wearable-computing/ (2013)
11. Chawla, N.V., Davis, D.A.: Bringing big data to personalized healthcare: A patient-centered framework. Journal of general internal medicine **28** (2013) 660–665
12. Chen, L., Nugent, C.D., Wang, H.: A knowledge-driven approach to activity recognition in smart homes. Knowledge and Data Engineering, IEEE Transactions on **24** (2012) 961–974
13. Provost, F., Fawcett, T.: Data science and its relationship to big data and data-driven decision making. Big Data **1** (2013) 51–59
14. Brule, M.R., et al.: Big data in exploration and production: Real-time adaptive analytics and data-flow architecture. In: SPE Digital Energy Conference, Society of Petroleum Engineers (2013)

15. Olston, C., Reed, B., Srivastava, U., Kumar, R., Tomkins, A.: Pig latin: a not-so-foreign language for data processing. In: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, ACM (2008) 1099–1110