Future Generation Computer Systems 56 (2016) 333–359

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Wiki-Health: From Quantified Self to Self-Understanding

Yang Li, Yike Guo *

Department of Computing, Imperial College London, United Kingdom

HIGHLIGHTS

- Propose a hybrid storage approach for managing heterogeneous formats of sensor data.
- Wiki-Health provides a unified solution for managing personal health sensor data.
- Wiki-Health Analysis Framework allows users to reuse and remix data and models.
- Analysis Tasks Allocation Scheme (ATAS) targets application level performance metric.
- Present an efficient ECG-based health monitoring service on top of Wiki-Health.

ARTICLE INFO

Article history:
Received 28 February 2015
Received in revised form
10 August 2015
Accepted 16 August 2015
Available online 24 August 2015

Keywords:
Wiki-Health
Big Data Management
Health sensor data management
Self-Understanding
Quantified-Self

ABSTRACT

Today, healthcare providers are experiencing explosive growth in data. Although the dramatic increase in the use of medical imaging technologies has been a major contributor to healthcare data growth in the past decade, more recently the rising adoption of sensing devices, enabling people to collect health-related data independently at any time or place is leading to a torrent of sensor data. The scale and richness of the sensor data currently being collected and analysed is rapidly growing. The key challenges that we will be facing are how to effectively manage and make use of this abundance of easily generated and diverse health data.

This paper explores the potential for sensors use in healthcare data acquisition and presents the next evolution in the on-going development of Wiki-Health, a big data service platform, designed to address the larger problem of explosive growth in healthcare information by providing a unified solution for collecting, storing, tagging, retrieving, searching and analysing personal health sensor data. Additionally, the platform is designed to allow users to reuse and remix data, along with analysis results and analysis models, to make health-related knowledge discovery more available to individual users – including health professionals, patients or even individuals who desire to maintain an optimum level of personal health – on a massive scale.

To tackle the challenge of efficiently managing the high volume and diversity of big data, Wiki-Health introduces a hybrid data storage model capable of storing structured, semi-structured and unstructured sensor data and sensor metadata separately. The design of such a hybrid model allows Wiki-Health to potentially handle heterogeneous formats of sensor data. In addition to its data management capabilities, we envision the potential for Wiki-Health as a system that also enables health sensor data monitoring and analysis, not only as a method of tracking existing health conditions but also as a means of encouraging a more pro-active approach to healthcare through early detection. To tackle the scalability and performance challenges of real-time analysis, the Analysis Tasks Allocation Scheme proposed in the research aids the management of data analysis tasks on a large scale and utilises the elastic nature of cloud infrastructure by considering the aspects of performance and cost.

To evaluate the proposed Wiki-Health approach, we have developed an ECG-based health monitoring service on top of the Wiki-Health platform. The positive performance of the approach is supported by the results obtained in our experimental trials and shows significant potential for real-world applications.

1. Introduction

Healthcare providers are constantly generating large volumes of data. According to industry estimates [1], worldwide healthcare

* Corresponding author.
E-mail address: y.guo@imperial.ac.uk (Y. Guo).

http://dx.doi.org/10.1016/j.future.2015.08.008
0167-739X/

data is expected to grow from 500 PB in 2012 to 25,000 PB by 2020. Such immense data ranges from complex, high-resolution medical imaging data, such as X-rays and CT scans, to the simple numerical data of systolic and diastolic pressure measurements from blood pressure devices. The means by which health care information can be gathered are also evolving. The growing global popularity of smartphones and tablets has led to the development of new ways of gathering information, both manually and automatically, by means of an array of embedded sensors. Professional wearable biosensors can connect to smart phones and track a significant number of physiological parameters. This combination of technologies has provided a more efficient and convenient way to collect a wide array of personal health information including blood pressure, oxygen saturation, blood glucose levels and pulse rates, as well as electrocardiogram (ECG), electroencephalogram (EEG) and electrocardiography (EKG) data. The different data formats used in the collection of healthcare data can be generally classified into three types: unstructured, structured and semi-structured. Examples of unstructured data include medical images, videos, plain text files and PDFs. Most types of structured sensor data generated from healthcare devices such as blood pressure and ECG, can be represented in the form of time-series.

With the introduction of increasingly sophisticated new technologies, the scale and richness of mobile sensor data being collected and analysed is rapidly growing. The use of sensor data is also an increasingly important aspect of the evolving healthcare landscape. Innovations in sensor monitoring can pave the way for a shift from the traditional reactive approach to healthcare – treating problems at the crisis level – to a proactive health management approach that allows health issues to be discovered and addressed at an early stage. Being able to monitor users' bio-signals over the long-term shows a great potential to allow us to understand their typical lifestyles and behaviours. More importantly, it allows for the tracking and discovery of any change-signals that could serve as early warning signs of potential health issues. Extensive research has been undertaken on various biological and physiological models in an effort to understand the means by which different physical and mental health conditions may be tracked and monitored. However, many of these models require inputs and parameters that still cannot be measured directly by sensors. Fusing and integrating data generated from different sensors requires considerable domain knowledge and experience.

As more people are becoming interested in gathering information about themselves by adopting these new technologies, monitoring their health and researching their own data, there is a clear need for an infrastructure on which scientists, developers, and professionals can publish their developed data analysis models as utilities, thereby enabling other users to access those services and utilise their collected sensor data without any expert knowledge. However, there are still many challenges to face. For example, it is still difficult for users to manage or utilise the data they collect for different purposes. From the provider's perspective, such massive growth in the volume of big health sensor data creates both data manageability and collaboration challenges. The emergence of cloud computing is seen as a remedy to these issues.

Cloud computing has been widely discussed in the past few years, as it shows great potential to shape the development, maintenance and distribution of both computing hardware and software resources. Within this computing paradigm, the actual provision of resources is only a concern at run-time for specific application requirements, as they can also be accessed in an on-demand and pay-per-use fashion. Cloud storage takes hardware sharing one step further: unlike local storage, cloud storage relieves end users of the task of upgrading their storage devices constantly. Implementing cloud computing technologies appropriately can aid healthcare providers in improving the quality of medical services and the efficiency of operations, sharing information, improving collaboration, and managing expenditures.

In our previous work [2], a cloud-based personal health sensor data management platform named Wiki-Health was presented, which is designed to tackle some of the challenges of storing, managing and analysing health sensor data. We envision Wiki-Health as having a role not only in alleviating the existing problem of medical information management, but also as a means of spearheading a shift in focus towards a more proactive model of healthcare management, one capable of helping individuals achieve healthier lifestyles and greater well-being through the adoption of the emerging technologies, such as smartphone applications and other devices, to independently track and monitor a variety of chronic health conditions. Wiki-Health is not only designed to solve the problems of managing and storing the increased volume, velocity and variety of data, but also to provide support tools for users to create healthcare-related applications and analysis models. As a result, the platform will allow them to effectively utilise all of the data generated from different data sources while reducing the complexity of dealing with its diversity. The functionalities and associated tools of Wiki-Health make it feasible for researchers to use as a test bed for undertaking health-related experiments and delivering novel analysis algorithms.

The principle aim of this paper is to present an overview of the most recent evolution of the Wiki-Health platform, which includes a newly developed data analysis framework, data visualisation tools and a sensor monitoring service. "Health monitoring as a service" is introduced as a key application for future personal health management. This is achieved through demonstration of the analysis model, and the design and implementation of an ECG-based personal health monitoring service application. The proposed Adaptive Learning Approach (ALA) within the analysis model aims to reduce the training time while showing improved performance over existing methods.

The organisation of this paper is as follows. In Section 2, we present an overview of the currently available technologies for healthcare data acquisition and a survey of sensor use for healthcare monitoring. This includes both established methods for monitoring cardiac activity and asthma inhaler use, as well as some proposed scenarios for monitoring other conditions. In Section 3, we present an overview of the design rationale of the Wiki-Health platform. In Section 4, we present the implementation details of several key components of Wiki-Health, including the Wiki-Health mobile app, with which users can interface with the platform, as well the graphical interface of Wiki-Health's Data Management Portal and visualisation tools to help users view their collected sensor data. This section also includes details of the Model Execution Engine for the platform. In Section 5 we evaluate the performance of the Wiki-Health platform for time-series based sensor data management under different workloads. Following this, a case study for enabling health monitoring as a service on top of the Wiki-Health platform for managing and analysing ECG signal data used in the diagnosis of cardiac problems is presented in Section 6. Conclusions drawn from the performance evaluations of the system's most recent features and a discussion of future work are presented in Sections 7 and 8.

2. Background and related work

In this section, we first present a background survey of the current health data acquisition modalities and challenges in terms of explosive data growth emerging from modern technologies such as professional bio-sensors and the pervasive use of sensor devices. Secondly, we explore the existing technologies as solutions for