A Cloud-based Mobile Data Analytics Framework

Case Study of Activity Recognition Using A Smartphone

Bingchuan Yuan
Department of Computer Science
University College Cork
Cork, Ireland
Email: by2@cs.ucc.ie

John Herbert
Department of Computer Science
University College Cork
Cork, Ireland
Email: j.herbert@cs.ucc.ie

***Abstract*—Unobtrusive gathering of personal or environmental data using a smartphone can provide the basis for intelligent assistive services. Continuous gathering of data will result in huge amounts of data, especially if many users are involved. Ideally, one might want to keep a large amount of this raw data for future (and maybe different) analysis, and also analyse the data to produce a compact model which can be used in the smartphone for real-time analysis of new data. This motivates a cloud computing solution where data from many users can be stored and analysed efficiently, and then the compact results of the analysis can be downloaded and used in the smartphone. This cloud-based approach is demonstrated using a case study of an activity monitoring application which might be used, for example, to monitor the daily activities, such as walking or going upstairs, of an at-risk person living alone. The cloud-based machine learning uses multiple classification methods, and, starting from individual training sets, enhances and builds classification models for each individual. The cloud-based system also builds a universal model based on all users which can be used as the initial classification model for a new user. The classification model produced by the cloud-based system is downloaded to the smartphone, and can be used to produce accurate real-time activity analysis. As more data is gathered and continually uploaded to the cloud, the models are adapted using an unsupervised learning approach to produce enhanced models which are then downloaded onto the smartphone for improved real-time activity analysis. The evaluation results indicate that the proposed approach can robustly identify activities across multiple individuals: using model adaptation the activity recognition achieves over 95% accuracy in a real usage evaluation.**

***Index Terms*—Cloud-based Data Analytics, Smartphone, Wearable Wireless Sensor, Machine Learning, Model Adaptation, Real-time Activity Recognition**

I. Introduction

Driven by quality and cost considerations, healthcare systems have to change radically in the near future from current healthcare professional-centric systems to distributed networked and mobile healthcare systems. In this change, a leading role will be played by pervasive technologies. More exactly, this approach is going to aid prevention and not replace traditional healthcare. Pervasive healthcare tries to change the healthcare delivery model: from doctor-centric to patient-centric, from acute reactive to continuous preventive, from sampling to monitoring. A pervasive system for healthcare helps a patient adjust his lifestyle to his health requirements.

Our previous work [1]–[3] describes *CARA* (Context-aware Real-time Assistant) an intelligent system especially designed for pervasive healthcare. The ability to accurately recognize and continuously monitor activities of daily living (ADL) is one of the key features that the CARA system is expected to provide. ADL is associated with both physical and mental health and is a primary indicator of quality of life [4]. Indeed, some age-related diseases(cognitive impairments, mild dementia and Parkinson's disease) have a direct impact on the ADL of the elderly [5]. In a previous paper, we introduced a novel and robust ADL recognition mobile application for pervasive healthcare [6]. This application identifies a user's activities through the combined use of a smartphone and a wearable bio-sensor belt. The solution involves a combination of a threshold-based technique for identifying simpler static activities, and a sophisticated machine learning technique for identifying more complex dynamic activities. It can be used to detect changes in a subject's routine and to predict symptoms over time through behavioural pattern recognition.

This work focus on introducing the development of a cloud-based data analysis framework which is designed to refine a classification model through multiple machine learning algorithms so that the model becomes customizable and adaptive for different individuals. Cloud computing is ideal for both storing the data and performing the compute-heavy machine learning. With cloud computing, the mobile device can go beyond simple recording of data. In this case, the cloud server provides an efficient means of data sharing and data mining which overcomes limitations of the mobile device.

II. State of The Art and Related Work

Advances in ubiquitous and pervasive computing have resulted in the development of a number of sensing technologies for capturing information related to human physical activities. Different approaches for lifestyle monitoring in terms of activity recognition have been studied using different underlying sensing mechanisms.

The environmental sensor-based approach has received significant attention in recent years. It is a promising approach for recognizing activities which cannot be simply distinguished by body movement alone. Additional sensors such as motion sensors, door contact sensors, pressure sensors, object RFID tags and video cameras are required for gathering activity related information [7]. However, it requires installation of a lot of equipment, and in some cases it is only feasible for use in laboratory settings. On the other hand, wearable sensors have proved to be an effective and reliable method for human activity recognition. They are small in size, lightweight, low cost and non-invasive. Some of the existing work on wearable sensor based activity recognition utilizes multiple accelerometers placed on different parts of the body [8], [9]. Other research has explored the use of multiple kinds of wearable sensors for activity recognition [10]. Nevertheless, current body-fixed sensors may be considered to be inconvenient to wear and impractical for continuous long term monitoring in a normal daily living environment. On the other hand smartphones have the potential to be an alternative platform for activity recognition with the advantages of unobtrusiveness and not requiring any additional equipment for data collecting or processing [11]–[13].

Using a smartphone as a primary device for data collection and processing increases the likelihood of data coverage and represents a minimal cost and maintenance commitment to the user. Although smartphones continue to provide more computation, memory, storage, sensing, and communication bandwidth, the smartphone is still a resource-limited device if complex signal processing and inference are required. Signal processing and machine learning algorithms can stress the resources of the phones in different ways. In particular, for activity recognition applications that require continuous sensing and real-time processing, it is very resource demanding [14], [15]. Fortunately, cloud computing can resolve the issue, for example, by utilizing a cloud with large storage capacity and powerful processing ability, the system can offer a cost-effective way to support data analysis technologies with high flexibility, scalability and availability for accessing data, discovering patterns, and deriving models.

There is a lot of undergoing research regarding mobile cloud computing, and a wide range of potential mobile cloud applications have appeared in the literature [16]–[18]. These applications fall into different areas such as image processing, natural language processing, multimedia search and sensor data management. For instance, [19] presents the benefits of combining machine learning techniques and cloud computing to enhance a course of image/video processing. Through mobile clients, users can understand and compare different algorithms processed in the cloud environment.

In this work, we focus on introducing a cloud-based data analytics framework especially designed to refine classification models for activity recognition utilizing multiple machine learning methods. By utilizing the cloud infrastructure, the smartphone, even with limited computational resources, can perform intelligent real-time classification and this provides novel functionality in our solution.

III. Cloud-based Data Analysis Architecture

The growth of sensors and intelligent systems has created a whole new warehouse of valuable data for healthcare. The issues raised by large data sets in the context-aware applications are becoming even more important. The problem in healthcare is never the lack of data, but the lack of information that can be used to support decision-making, planning and strategy. By applying effective data analysis technology to the data gathered from sensors and intelligent systems, it could make decision support for healthcare more efficient and ultimately more accurate.

Data analysis technology is designed to work with large volumes of heterogeneous data. It uses sophisticated statistical methods such as machine learning, computational mathematics, and artificial intelligence to explore the data and to discover interrelationships and patterns. Deeper insights are possible when more data is available. As a delivery model for IT services, cloud computing can enhance the potential of scalable data analysis solutions while enabling greater efficiencies and reducing costs. As the cloud infrastructure is distributed and fault tolerant, a cloud-based data analysis framework can be deployed on pools of server, storage, and networking resources that can be scaled up or down as needed. Since servers are virtualized, different instances can reside on the same hardware. The instances can be moved around depending on the need to make the best use of the hardware without compromising performance. Indeed, cloud computing offers a cost-effective way to support data analysis technologies and intelligent healthcare applications with high scalability, availability and computing power for accessing healthcare data, discovering patterns, and deriving models.

In the CARA system, a comprehensive cloud-based data analysis framework is developed by combining big data analytics and cloud computing technologies. This provides analysis as a service - from data delivery and data analysis to data storage, in order to optimize the total value of healthcare data. A general structure of the cloud-based data analysis framework is presented in Fig. 1.

IV. Case Study of Activity Recognition

In the case study of CARA activity recognition, a custom-design Android application is implemented and run on Samsung Galaxy III for real-time classification. A private cloud server is deployed in Window Azure providing Software as a service (SaaS) for activity data analysis. The process of control flow is shown as follows:

1) Machine learning classification model is loaded from the cloud into a smartphone.
2) Raw data are collected from BioHarness sensors and smartphone built-in accelerometer and gyroscope sensors.
3) The signals are segmented and a 1s window is moved over the signal and overlapped every 500ms.
4) The features corresponding to each window are extracted.
5) Static and dynamic activities are distinguished using a threshold-based method where a threshold value of the acceleration is applied.

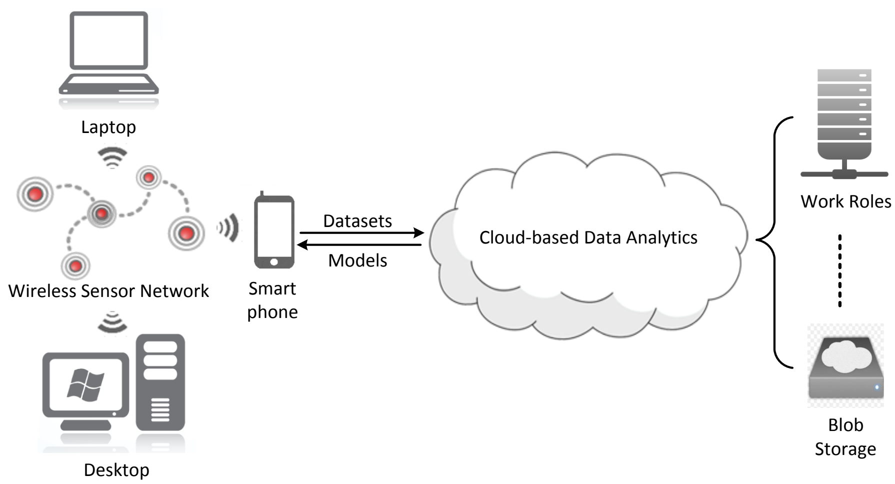


Fig. 1. Structure of the activity recognition system

6) Static activities are divided into sitting, standing, bending, lying and leaning back by applying threshold angles for both trunk and thigh.
7) Dynamic activities are classified using machine learning classifier based on the dataset of extracted features.
8) Activities are labelled as correctly classified if the output confidence is high enough.
9) Features and the activity label of each detected case are stored in the data file.
10) Recorded data are used to retrain the classifier and refine the classification model using the cloud-based data analysis framework.

Wearable wireless sensors and smartphone sensors generate huge amounts of raw data while the patient is under monitoring. Data are transmitted to the smartphone in real-time. Features of the data are then extracted and built into datasets which contain valuable information and the context required by the reasoning engine. However, the data without a model is just noise, they need to be further processed and built into models in order to be useful for activity recognition and decision making. Models are used to describe salient features in the dataset, which can be constructed via machine learning methods. In this work, we investigate different machine learning algorithms to build classification models for activity recognition (i.e. Decision Tree, Neural Network, Nearest Neighbour and Bayesian Network), each of which can be viewed as a blackbox that is capable of inferring current situation based on the input dataset. However, each model comes from a unique algorithm approach and will perform differently for the same dataset. The best approach is to use cross-validation to determine which model performs best for a given dataset. To achieve that, we deploy several machine learning worker roles on a private cloud server which share the blob storage. Each worker role holds an instance of one type of machine learning method for dealing with the input dataset. Once the new dataset is uploaded to the cloud by the user, the worker roles start analysing data in parallel, producing machine learning models as well as evaluating them. After examining the evaluation results, the best model is selected to support activity recognition in the CARA system. Moreover, models and datasets are stored in the blob as resources to be shared with other applications or healthcare systems. Further training datasets would be generated locally in a home environment through using the best classification model and can be uploaded to the cloud server periodically to get update for new models. Thus, it provides an unsupervised learning mechanism for the CARA system, and consequentially it becomes more accurate and reliable.

A. Data Collection

The data were collected by using a Samsung Galaxy III mobile phone [20] and wearable Zephyr BioHarness sensor as shown in Fig. 2 [21]. The embedded tri-axial accelerometer and gyroscope sensors measure the 3D-acceleration and orientation of the smartphone. The three axes of acceleration are dependent upon the orientation of the phone, the x-axis runs parallel to the width of the phone, the y-axis runs the length of the phone, and the z-axis runs perpendicular to the face of the phone.

The sensor integrated into the mobile device is easy to use without assistance and can be carried comfortably for long periods of time [22]. The data collection was done by performing experiments on eight postgraduate students at University College Cork, Ireland. Subjects were asked to perform a series of activities while carrying the smartphone in the front pocket of their trousers and the BioHarness sensor on the chest. The BioHarness sensor data were transmitted to the smartphone through a Bluetooth connection in real-time. The real-time sensor readings were recorded for the purpose of training the machine learning classification models.

Fig. 2. Body area network for sensor data collection

B. Feature Extraction

After obtaining sensor readings, the machine learning classifier needs to be trained to build statistical classification models for real-time activity recognition. However, standard classifiers may not work well on the raw sensor data due to the characteristics of sensor readings, e.g. instability and noise. It is essential to transform the raw sensor readings into a dataset that captures the significant features of the raw data. This is usually performed by breaking the continuous data into windows of a certain duration. In this work, we experimented with one second time window which is overlapped by one half of the window length. Hence, each window is a single instance, but any given data point contributes to two instances. This method has been shown to be effective in earlier work using accelerometer data [23]. We compute both time-domain and frequency-domain features for each axis of accelerometer and gyroscope readings. The time-domain features measure the temporal variation of a signal, and consist of following four features. The dynamic range is defined as *Min*, *Max* and *Mean*, which represents the minimum, maximum and mean value in a time interval. The *Standard Deviation* is calculated to characterize the stability of a signal, normalized by the mean value of the readings in the interval. For the frequency domain, we consider two features. The *Zero-Crossing Rate* and *Mean-Crossing Rate* indicate the frequency of sign-changes along a signal in a time interval, which measures the rate of signal changes from positive to negative and from higher than the mean value to lower than the mean value. Thus, a total 66-dimensional feature vector is generated every second. The extracted features actually generate a pattern for certain activities, and these patterns, consisting of feature vectors, are then used to build machine learning classification models.

C. Real-time Classification

To recognize dynamic activities such as walking, walking stairs and jogging, the machine learning classifier was trained to produce the classification model. The Weka machine learning package [25] was used in this study for developing the machine learning mechanism for the activity recognition in real-time. This package provides a collection of machine learning algorithms for data mining tasks which can be used in the android application. Five general machine learning algorithms were investigated and evaluated in this work: *Bayesian Network, Decision Tree, K-Nearest Neighbours, Support Vector Machine and Neural Network*. We used the default parameters associated with each of the classifiers. The classifier obtains a model during the training stage. After training, it can predict the class membership for new instances using the model.

V. CLOUD-BASED MODEL ADAPTATION

At the beginning of our experiment, data collected from each individual are manually labelled for supervised learning. The individual's data are randomized and separated into a training set and testing set. The classification model is built on the training set and evaluated on the testing set for each individual, which is referred to as the *Personalized Model*. Although the *Personalized Model* is able to provide a better result for a specific user, it may not be reliable or suitable for other users. In other words, it has the limitation of usability and scalability. Hence, a generalized model is required to cater for all users. Data collected from all the subjects are pooled together to build a universal classification model referred to as the *Universal Model*. It works as the default for a new user when there is no existing personal model available for that user. We evaluate the *Universal Model* for each individual using different machine learning classifiers. However, it turns out that the performance of different classifiers varies from person to person which mostly depends on the training dataset collected from each individual. In other words, it may not exactly fit each individual regarding a specific activity because of physical differences between individuals. In addition, the supervised learning scheme requires a great effort of manually labelling the activity data which is a tedious and a heavy task for the user. The process of training and evaluating models consumes a large amount of time and computational resources, and this is difficult to do on a smartphone with limited computing power while the real-time activity classification is also proceeding. In order to improve the model accuracy as well as to enhance the classification performance, we introduced the idea of model adaptation for the optimization of the machine learning classifier by utilizing the cloud infrastructure.

As shown in Fig. 3, the principle of model optimization is to keep updating the classification model for an individual user while the activity recognition task is carrying on. All the new users start with a default classification model (*Universal Model*), which in turn gets refined and adapted to each individual user for better performance when more activity data is available as users carry the phone.

On the client side, real-time activity recognition is carried out in the smartphone using the classification model. An unsupervised learning scheme was applied to generate new training data which is based on a self-training technique using

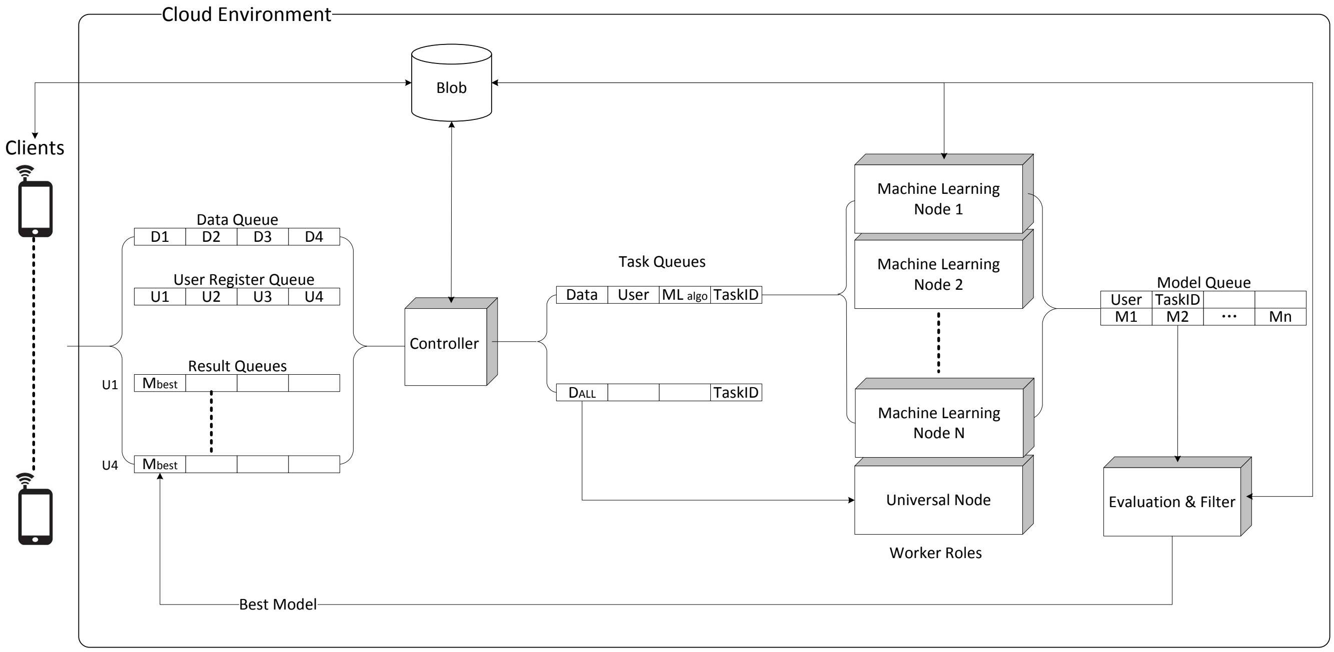


Fig. 4. Structure of cloud-based data analysis framework

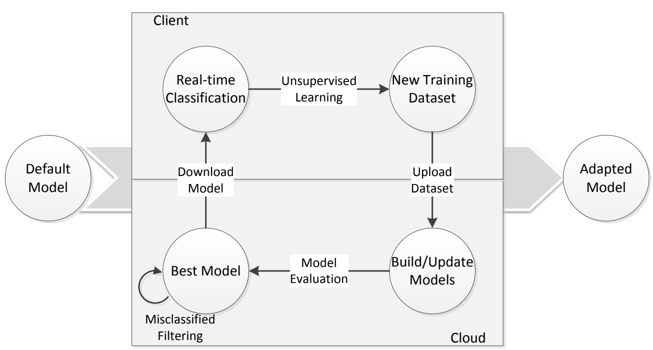


Fig. 3. The process of model optimization

unlabelled data without any user input. It reuses the predicted label and confidence statistics generated by the machine learning classifier during the inference process to select new training samples. The adaptation method determines whether a data sample is suitable for adaptation according to the confidence level of the classification result and uses only high confidence samples. If the normalized likelihoods of the two classes are quite close to each other, the classifier will have a low confidence on the prediction. Hence, a confidence threshold is used to filter the equivocal samples. Because unlabelled data is cheap and abundant, the system can use a very high threshold to ensure data quality. The new training data for adaptation are stored in the smartphone temporarily and are uploaded to the cloud storage periodically.

On the cloud server, blob storage, multiple queues and worker roles are deployed for the purpose of analysing a user's activity data and producing the best classification model for each user. The blob storage of the cloud stores all the available activity data and classification models separately for each user. Each single user is assigned an independent container which contains individual activity data, including existing data and new uploaded training data, and the adapted classification model for this user.

The framework consists of five types of queues:

- **Data Queue**: This queue is a general queue and is used to communicate between the Client and the Controller Node. When a client uploads a training file into the Blob the URL of the link plus the user id is added in to the Data Queue.
- **Result Queue**: When the most suitable model is selected by the allocated worker role (Evaluation Node), the URL of the best model is added to that user's Result Queue. Unlike the Data Queue which is unique and all the users have access to it, the Result Queue is independently created for each individual and doesn't provide public access.
- **Register Queue**: It is used to transfer the information of a new user. Once a new user is registered, the system assigns him/her a dedicated Result Queue and a blob container. The default model will be sent to the new user initially.
- **Task Queue**: In order to make all worker roles start building models in parallel, there should be separate queues. The queues are dedicated to each of the worker

roles. The Controller reads data from the Data Queue and assigns it into different Task Queues. There is a separate queue for the Universal Node which will update the default model based on the data from all users.
- **Model Queue**: Once the model of each classifier is retrained and evaluated, the evaluation result will be sent to the Evaluation Node in the Model Queue.

The definition of each worker role and their responsibilities are listed below:

- **Controller Node**: The main role of the controller is to control and manage the flow of incoming data and add the new messages to the Task Queue for further processing.
- **Machine Learning Nodes**: Each of the worker roles reads a message from its own Task Queue and starts producing a model based on a different machine learning classifier. When the model is retrained, it will be evaluated by using a manual labelled testing set.
- **Universal Node**: Unlike the other machine learning worker roles, it is designed to deal with data from all the users to update the Default Model using a dedicated classifier (e.g. Neural Network). The model is evaluated using a cross-validation method and the misclassified instances are filtered.
- **Evaluation Node**: The duty of this worker role is to select the most suitable model for a specific user. It can be achieved by comparing the evaluation results (e.g. error rate, precision, recall) of each classification model (excluding the universal model). Moreover, a misclassified filter is also applied to the user's data after the best model is found.

The process of cloud-based model optimization is illustrated in Fig. 4. Once new training data is uploaded by the client, the URL of the training data is collected by the Controller node and assigned as different tasks to machine learning worker roles. The data are then used to either build new classification models in case of a new user or refine the existing models. Moreover, the data is also fused with the available data from all users to update the *Default Model* in the universal node. In the deployment, multiple machine learning nodes work in parallel to process the data analysis tasks. Each of the worker roles is designed to handle one popular machine learning algorithm. The number of worker roles can be adjusted on demand which makes use of the flexibility and scalability of the cloud infrastructure. The models are generated and evaluated for each machine learning classifier in worker roles. After gathering the evaluation results of all the models in the evaluation node, the most suitable model is selected according to the performance of each model. The data (combining new training data and any previously existing data) for that user are then filtered by a misclassified filter using the best model and stored in the cloud storage as previous data. Once the process is complete, the best model is sent back to the user automatically. Thus, the new or updated model can be used in the client for real-time classification.

One of the advantages of the cloud-based data analysis framework is that with a queue-based architecture an asynchronous scheme is applied where worker roles are asynchronously coupled which means that scaling or adding/removing instances does not affect the other worker roles. In this case, there is no hard dependency between cloud nodes. The framework provides horizontal ***scalability*** for the various machine learning tasks. Furthermore, [26] pointed out that the size of the classifier model is independent of the size of the training data whereby even though the training data set is very huge the model does not need to be very big. Huge amount of training data can be stored in the cloud while models generated from the training data can be downloaded and used in any mobile device. That can be interpreted as another advantage of this framework: improving the ***availability*** of the smartphone-based activity recognition system. According to our experiment, the training data with size of 54.4 MB, which contains 1 hour activity data (more than 10000 samples), only generates a model file with size of 170 KB (Decision Tree classifier). Last but not least, the framework also minimizes the processing time of the machine learning tasks by leveraging the ***computing power*** of cloud infrastructure which enables a quick response in the mobile client.

VI. Experiments and Results

The experiments were conducted in a home setting. Activity data was collected from eight volunteers (five male, three female) and manually labelled for training and testing the machine learning classifiers. Two more subjects (one male, one female) were involved in the second experiment to help us evaluate the feasibility of the proposed cloud-based model adaptation approach. To collect data for supervised learning, subjects were asked to perform static and dynamic activities sequentially while carrying the smartphone and wearing the body sensor. A controller application running on an Android tablet remotely controlled the smartphone to start or stop recording activity data through bluetooth connection and to label the recorded data on the fly. The recorded activity data was used to build the default model in a supervised learning manner, it also worked as the gold standard testing set to evaluate the adapted classification model generated for each individual later.

In the first experiment, we build the *Personalized Model* which is a completely user dependent approach. It requires training machine learning classifiers on each individual user's activity data and generates a user dependent model for each user. Clearly, this scheme is superior to other models in terms of performance for a specific user, but its lack of general usability and scalability greatly limits its application. In addition, the supervised learning scheme requires a great effort of manually labelling the activity data which is a tedious and heavy task for the user. To improve it, we decided to move on to a semi-supervised approach where a universal model is trained with manually labelled data from all users (called *Default Model*). As we discussed in the last section, an unsupervised learning scheme was applied to gather new training data to update the model so that it gets adapted to each individual user in a progressive manner. In this case, an *Adapted Model* was gradually generated for each user. We investigated five popular machine learning classifiers in the local trial run and evaluated the real-time activity recognition approach on 8 subjects respectively. In a case study of the Bayesian Network classifier, the accuracy for each of the subject with different models is shown in Figure 5. As shown in the plots the universal model is penalized for its one size fits all philosophy. The *Universal Model* provides the lowest accuracy of 80.17%. The *Personalized Model* provides the highest accuracy of 99.19% and the the *Adapted Model* yields 93.16% average accuracy.

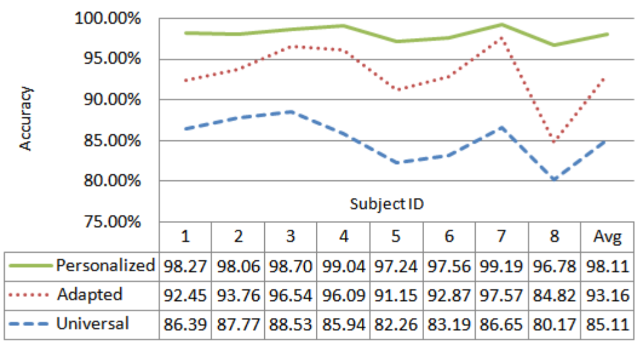


Fig. 5. Accuracy of different models using Bayesian Network classifier

The proposed cloud-based approach was tested on two users (one female user A and one male user B). The manually labelled data gathered from these two users were used as the testing set to evaluate five classification models. Each user started with the *Default Model* and was able to obtain his/her *Adapted Model* after the first run. The model was refined and updated while the experiment was carried on. The impact of the model adaptation mechanism can be seen in Fig. 6 which illustrates the activity recognition rate with various machine learning classifiers. After four runs of each user, the overall accuracy of the classification model was boosted and the best performance achieved over 98%. It can be observed that the performance of each model in the first run is quite poor, because the *Default Model* doesn't match the user very well. However, after updating the *Adapted Model* and filtering the misclassified instances in the next few runs, the accuracies of most models remain consistently above 90% except for K-NN classifier for user A. The best accuracy was obtained with the Neural Network method while the most cost efficient model was the Decision Tree. Note that the execution time for building and evaluating the classification model using each machine learning algorithm is listed in the parentheses. It only takes about 1.5s to execute the Decision Tree algorithm while the processing time of the Neural Network linearly increases with the number of instances. In this case, the Decision Tree was considered as the best *Adapted Model* for this user. The *Adapted Model* represents a middle ground between the *Default Model* and the *Personalized Model*; it got refined and eventually yielded 98% accuracy after four runs through model optimization, and it is the model that is actually used in real-time activity recognition.

In addition, we measure the correlation of the different features using information gain based on the attribute ranker search method [25] which evaluates the worth of an attribute by measuring the correlation between it and the class. We also evaluate the worth of a subset of features using the greedy step-wise method by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. Table I presents the top 10 features evaluated on all user data using both methods. It is clear that the smartphone's accelerometer and gyroscope reading are ranked as the most important signals, and standard deviations of signals are selected as the most predictive features.

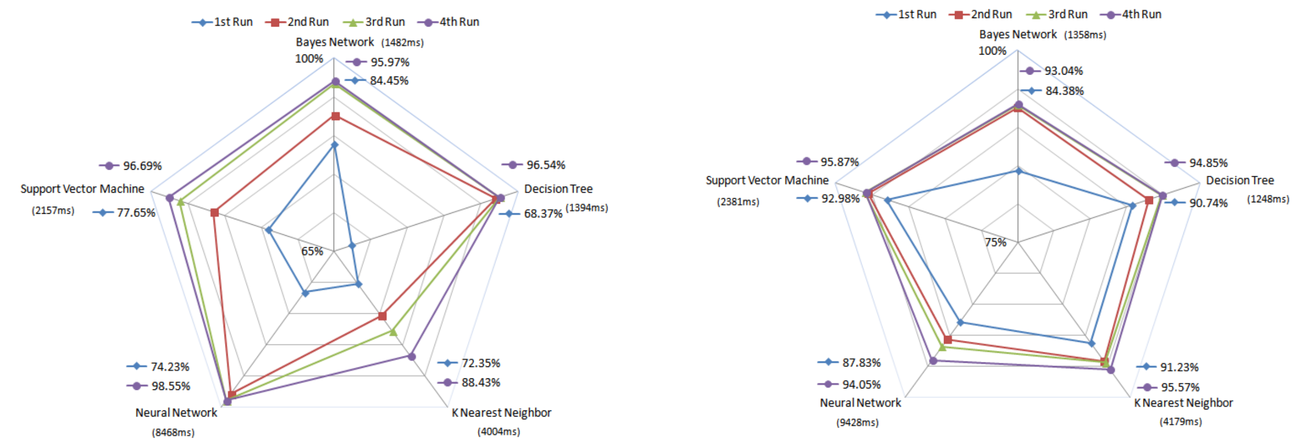
TABLE I
FEATURE RANKING EVALUATED BY TWO DIFFERENT METHODS

Rank	Correlation	Subset
1	AbsGyoMax	AbsAccMax
2	AbsGyoMean	AbsAccMin
3	GyoX Standard Deviation	AbsAccMean
4	AbsGyo Standard Deviation	AbsAcc Standard Deviation
5	GyoXMax	AccXMean
6	GyoY Standard Deviation	AccX Standard Deviation
7	AccY Standard Deviation	AccXZeroCross
8	AccZ Standard Deviation	AccYMin
9	AbsAcc Standard Deviation	AccY Standard Deviation
10	GyoXMin	AccZMin

VII. Conclusions and Future Works

Previously, we have developed a reasoning framework for the CARA pervasive healthcare [1]–[3]. This can help extend independent living for the elderly in a smart home environment by monitoring the person and ambient changes to detect anomaly situations [3]. For a smart home based monitoring system, it is important to detect human body movement which can provide us with basic activity context. By combining the activity context along with other environmental contexts, the context-aware monitoring system is able to perform better reasoning for healthcare.

In this paper, the daily activity monitoring using hierarchical classification is supported by a cloud-based solution. With integration of the cloud infrastructure, the system provides superior scalability and availability for data analysis and model management. The data processing and classification algorithms are implemented in the smartphone for real-time activity monitoring while the data analysis and model evaluation are conducted remotely in the cloud. The experimental results shows a significant improvement in comparison to the approach using a single smartphone [28] and the approach built on fixed machine learning algorithms [29]. This shows a lot of promise for using smartphones as an alternative to dedicated sensors and using the cloud-based data analytics framework to process machine learning tasks.



(a) Overall model accuracy for the female user A

(b) Overall model accuracy for the male user B

Fig. 6. Cloud-based model optimization for different classifiers

A limitation of this approach is that the number of features extracted from raw sensor data for machine learning is fixed in this work. In future work, it is necessary to assess the value of different features during the learning process so that the machine learning mechanism can be improved by assigning weights to features or by dynamically adjusting the number of features involved.

REFERENCES

[1] B. Yuan and J. Herbert, "Web-based real-time remote monitoring for pervasive healthcare," in *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2011.

[2] ——, "Non-intrusive movement detection in cara pervasive healthcare application," in *The 2011 International Conference on Wireless Networks*, 2011.

[3] ——, "Fuzzy cara - a fuzzy-based context reasoning system for pervasive healthcare," in *Procedia Computer Science*, vol. 10, 2012.

[4] D. White, R. Wagenaar, and T. Ellis, "Monitoring activity in individuals with parkinsonâĂŹs disease: A validity study," pp. 12–21, 2001.

[5] R. W. D.K. White and T. Ellis, "Monitoring activity in individuals with parkinson's disease: A validity study," *Jurnal of Neurological Physical Therapy*, vol. 30, 2001.

[6] B. Yuan and J. Herbert, "Smartphone-based activity recognition using hybrid classifier," in *In proceeding of the 4th International Conference on Pervasive and Embedded Computing and Communication Systems*, 2014.

[7] S. I. E.M. Tapia and K. Larson, "Activity recognition in the home setting using simple and ubiquitous sensors," in *Proceedings of PERVASIVE*, 2004.

[8] W. H. K. L. E.M. Tapia, S.S. Intille and R. Friedman, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor," in *Proceedings of the 11th IEEE International Symposium on Wearable Computers & Physics*, 2007.

[9] C. J. N.C Krishnan, D. Colbry and S. Panchanathan, "Real time human activity recognition using tri-axial accelerometers," in *Sensors, Signal and Information Precessing Workshop*, 2008.

[10] D. S. U. Maurer, A. Samilagic and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *IEEE Proceedings on the International Workshop on Wearable and Implantable Sensor Networks*, 2006.

[11] G. W. J.R. Kwapisz and S. Moore, "Activity recognition using cell phone accelerometers," in *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data*, 2010.

[12] S. L. A.M. Khan, Y.K. Lee and T. Kim, "Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis," in *5th International Conference on Future Information Technology*, 2010.

[13] N. K. B. T. D. Stefan, B. Das and D. Cook, "Simple and complex activity recognition through smart phones," in *8th International Conference on Intelligent Environments*, 2012.

[14] H. Gao and Y. Zhai, "System design of cloud computing based on mobile learning," in *Proceedings of the 3rd International Symposium on Knowledge Acquisition and Modeling*, 2010.

[15] X. Chen, J. Liu, J. Han, and H. Xu, "Primary exploration of mobile learning mode under a cloud computing environment," in *Proceedings of the International Conference on E-Health Networking, Digital Ecosystems and Technologies*, vol. 2, 2010.

[16] G. Huerta-Canepa and D. Lee, "A virtual cloud computing provider for mobile devices," in *Proceeding of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond*, 2010.

[17] R. Frederking and R. Brown, "The pangloss-lite machine translation system," in *Proceedings of the Second Conference of the Association for Machine Translation in the Americas*, 2010.

[18] E. Marinelli, "Hyrax: cloud computing on mobile devices using mapreduce," Master's thesis, Carnegie Mellon University, 2009.

[19] R. Ferzli and I. Khalife, "Mobile cloud computing educational tool for image/video processing algorithms," in *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop*, 2011.

[20] Samsung, "Android smartphone: Samsung galaxy siii," description Available online at http://en.wikipedia.org/wiki/Samsung_Galaxy_S_III.

[21] Zephyr, "Wireless professional heart rate monitor & physiological monitor with bluetooth," description Available online at http://www.zephyr-technology.com/products/bioharness-3/.

[22] T. Choudhury and S. Consolvo, "The mobile sensing platform: An embedded activity recognition system," *IEEE2008 Pervasive Computing*, 2008.

[23] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," *Pervasive Computing*, 2004.

[24] D. H. P. G. G.M. Lyonsa, K.M. Culhanea and D. Lyonsb, "A description of an accelerometer-based mobility monitoring technique," *Medical Engineering & Physics*, 2005.

[25] E. F. H. Ian and A. Hall, *Data Mining: Practical machine learning tools and techniques*, 3rd ed. San Francisco: Morgan Kaufmann, 2011.

[26] L. S. Kmiecik, "Cloud centered, smartphone based long-term human activity recognition solution," Department of Computing, Imperial College London, Tech. Rep., 2013.

[27] D. Cook, "Learning setting-generalized activity models for smart spaces," in *IEEE Intelligent Systems*, 2012.

[28] C. N. S. Zhang, P. McCullagh and H. Zheng, "Activity monitoring using a smart phone's accelerometer with hierarchical classification," in *6th International Conference on Intelligent Environments*, 2010.

[29] J. Andreu and P. Angelov, "An evolving machine learning method for human activity recognition systems," *Journal of Ambient Intelligence and Humanized Computing*, 2013.