HARU:Human activity and schedule recognition using sensors and GPS on smartphones

Wei Zheng [1] **Qi He** [1] **Shengyang Sun** [1]

Abstract

Equipped with many sensors, mobile phones bring much convenience to people's daily life. We propose HARU, which performs accurate activity recognition based on continuous collected accelerometer data, integrates the results with GPS data, and visually displays users daily life track. With real-time monitoring, HARU enables plentiful possibilities to live in a more intelligent way such as site recommendation and calorie consumption summary.

1. INTRODUCTION

These days, smartphones are equipped with a rich set of sensors. To name some sensors, camera collects images information, microphones receive sound around, accelerometer and gyroscope track the moving features of a phone, and GPS gets user's location precisely. These devices have become a part of our daily life, because we carry our smartphones nearly everywhere we go. There is a trend that people are more and more attached to the usage of smartphones (Lee et al., 2014). Therefore, with various sensors encapsulated, smartphones are enabled to track and monitor people's daily activities. There already are many applications on this. For example, Wechat Fitness counts user's daily walking steps, and Google Map can detect user's heading detections. As a result, sensors on the smartphone provide us with much more possibilities to explore and influence people's life.

There are already many works on activity recognition,(Anguita et al., 2012a) proposes to detect user status using SVM and shows high accuracy. (Martin et al., 2017a) compares many different algorithms including PCA, kNN, and random forests. In final random forest proves to have best recognition result. However, most previous works mainly focus on the accuracy of classification instead of moving on and explore its usage. In our work, we tend to combine the activity recognition and location information to provide a high level activity recognition and user's habit prediction. For example, if the user staying still (sitting) in a restaurant for one hour, we could predict that he/she is having lunch there. In addition, the system can provide information summary to users, such as daily schedule, time spent in transportation, energy consumption and life range, thus supply a reference for people's living arrangements.

In this paper, we mainly present HARU, a system that used to record individual daily life and predict the user's habit based on the sensor data on smartphones. The word "HARU" means spring and daytime in Japanese and "one day" in Korean. We would like to consider this project as a new attempt in the field of human activity recognition and human track monitoring. We believe this kind of data is very meaningful for user research, human life pattern study and helping users make a better plan and daily schedule. In the following sections of this paper we would describe the prototype implementation of our project about the architecture, sensor data collection, activity recognition, map rendering and information aggregation. Beside, we also evaluated our system with the ground truth data by using a number of different machine learning algorithms. In the end of this paper, we present the related work and future work.

2. IMPLEMENTATION

2.1. System Architecture

We first present the architecture design for HARU Fig 1 Smartphones are the data provider and consumer that HARU runs on the phone to collect accelerometer and GPS data from the user's mobile devices. The data collected will be aggregated and upload to the cloud for further processing. Two tables are created in AWS DynamoDB to store the raw accelerometer and GPS data. We developed GPS data preprocess unit and accelerometer preprocess unit on the cloud server to clean and preprocss. Accelerometer preprocessing module clean the accelerometer data and generate feature based on the window size. Raw data will be preprocessed every hour, and server can support real-time processing. The raw data would be deleted after one week of preprocessing.

[*]Equal contribution [1]Department of Computer Science, University of Toronto. Correspondence to: Wei Zheng <zhengw14@cs.toronto.edu>, Qi He <heqi4@cs.toronto.edu>, Shengyang Sun <ssy@cs.toronto.edu>.

Cleaned accelerometer data will be sent to action classifier. Classified actions with timestamps will be transfered to the aggregator. The aggregator takes classified action data and cleaned GPS data and produce the final results. The final output will be stored in a new table and user can access it from a webpage.

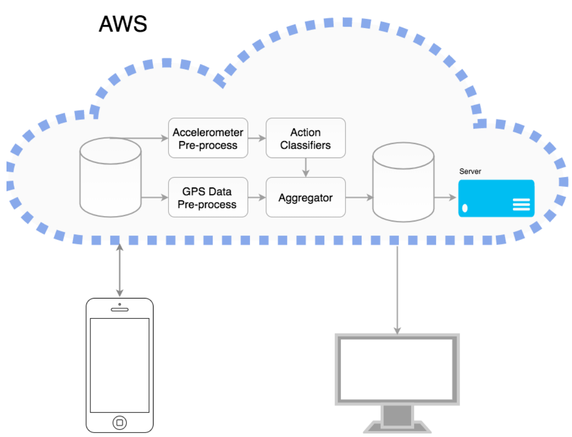


Figure 1. Architecture of HARU

2.2. Sensor Data Collection

In our project, we provided a iOS 11 application running on an iPhone 7 and an iPhone X to collect the accelerometer and location data. This accelerometer data includes the acceleration along the x-axis, y-axis and z-axis and the timestamp. These axes capture the horizontal/sideway movement of the user (x-axis), upward/downward movement (y-axis), and forward/backward movement (z-axis)(Bayat et al., 2014). Fig 2 demonstrates these axes relative to a user. The sample rate of the accelerometer data was set at 50HZ. Impact of window size on accelerometer data is disccuss in (Banos et al., 2014). For the location data, the sample rate was 1HZ because we do not need the high frequency and accuracy for this data. Besides, according to the normal moving speed, people's moving distance is limited within 1 second. The location data includes the features of longitude, latitude and timestamp.

In order to extend the capabilities of mobile devices we offload the computing tasks onto a cloud. Therefore, the data collected will be transferred to the database on the cloud for further processing. We use the AWS DynamoDB as our database. In terms of the uploading frequency, we tried different strategies. At first, we planned to upload the sensor data we collected in real time. Because of the constraints of mobile devices' battery life, network resources and high frequency upload request for the database, we abandoned this solution. Second, we designed to upload the data by the rate of every ten minutes. This solution was also be abandoned, because the size of data transferred once is up to more than 30000 items, which can not be supported by the batch storage method of DynamoDB or we need to pay a high price for this operation. Finally, we decided to transfer the data every minute based on the tradeoff between the power consumption and the database storage throughput.

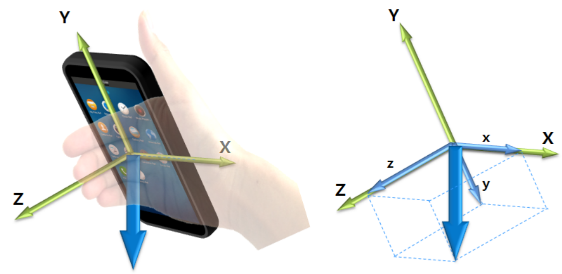


Figure 2. The axes of acceleration to user

Figure 3. The mobile user interface of HARU

To measure the accuracy of the activity classification, we provide a user interface developed with Swift language to enable the user to report their actual activities. The user interface is shown in Fig 3. The activity choices are designed as switch buttons, which are distinguished by different colours.There are five types of activities are involved in our applications, including staying still, walking, running, going up the stairs, going down the stairs. When users start to use our application, they need to turn on a certain button and the button color will change to grey. Then the application would start to upload the actual activity code and sensor data.

2.3. Activity Recognition

The core of HARU lies in the recognition techniques used for classifying activity. In our work, the data collected for activity recognition is the accelerometer data from mobile phone for three axes. Previous work (Micucci et al., 2017) has applied classifying algorithms on accelerometer data for classifying several different activities and different falling status and shows accurate recognition results. Based on their results, we also use accelerometer information without any information from other sensors like gyroscope. The saving of used sensors brings savings of energy, which is one of most important factors for mobile devices.

However, its nontrivial for accurate activity recognition based only on the accelerometer data for several factors.

1. Accelerometer consists of continuous data from three dimensions x, y, z. Because the data measured at one moment cannot reveal the movement feature, data must be collected successively and frequently, which brings heavy computing burden for the mobile phone.

2. Accelerometer data depends on the status of phones heavily. Even for the static case that there is no movements of the phone. Three axis data differs a lot according to whether the phone is placed flat or vertically, because of the influence of gravity. While in real life, the phone can be placed in the pocket or hold in hand, making its status differ a lot.

3. Accelerometer data information exists in the dynamic, thus the data can be very dependent and complicated. Specific classification algorithms must be chosen delicately to get a feasible activity recognition results.

For the high computing problem, by offloading the computation to the cloud, we can save large amount of energy cost. Gravity influence for three dimensions can be a big problem for us, as it interferes the true moving dynamic for activity recognition. Thus we propose to use beforehand data cleaning to remove it's signal.

We observe that in moderate length of time, the status of phone will not change a lot. For example, if the phone is placed in the pocket, then the phone will keep vertically whenever it is still in the pocket, only minor changes happen because of the movement of a person. Therefore, considering the frequency domain, low frequency regions carries most information for phone status and high frequency regions carries the information that how the phone or person is moving. Thus a high frequency passing filter will filter only the useful information for activity recognition.

Based on the observation for different frequency regions, we use an order 6 Butterworth filter to perform data cleaning on raw data and then applying classification algorithms.

2.4. Information Aggregation

Information Aggregator is the key module of HARU which is responsible for integrating the classified action data and GPS location data. GPS data and classified action data are generated using different frequency so information aggregator will need to combine them. Each of the record in GPS and classified action data has a timestamp. The integrator will take each GPS data entry's timestamp and find the record in the classified action data with the closest time to it. The classified action with the closest timestamp will be added to the GPS record. Consecutive GPS data with the same classified actions will form a group of GPS records with a classified action. A small GPS record groups will be added to the previous group of records. Less than 30 seconds of staying still groups, less than 10 seconds of walking and running groups are considered to be small. By merging the small groups with previous groups, we remove the noise. For example, the user could be waiting for the green light to cross the street; we do not want to classify that as staying still. For staying still groups, we use Google Place API to get the place information. By using the features, we get from Google Place API with the duration and start time of staying still we can predict what use is doing at the time. For example, staying still in a library in the afternoon means studying. Staying still in a restaurant means having lunch or dinner. The machine learning model we use to predict user behaviours is the decision tree. We do not have any data regarding user behaviour prediction, so we built the decision tree based on our knowledge. The GPS record groups with classified action will store in a table, and we could visualize it using Google Map.

2.5. Map Rendering & Schedule View

We use Google Map API to provide visual interface for HARU Fig 4. On Fig 4 we have an interactive Google map on the left and a scrollable list of cards on the right to provide a schedule view. For the interactive Google map Fig 5, HARU draw purple lines to represent walking path and orange line to represent running path. Center of the blue circles represent the place where the user stays still and larger the circle means longer the user stay in the area. Each of the card Fig 6 on the right represent certain action performed in a time interval. Each card Fig 6 will have a static picture generated using Google Map API showing the location or path information of the action. Currently we only show walking, running and staying still as going up stair and going down stair only account for small portion of the actions. The card will give start time and amount the time the action was performed. In addition to the time information, when the users is staying still we provide the prediction what the user was doing at the time. The cards are expandable; more information will be provided like nearby location and the city Fig 7. We plan to add more information and functionalities to the interface, and we will mention those in the future works section.

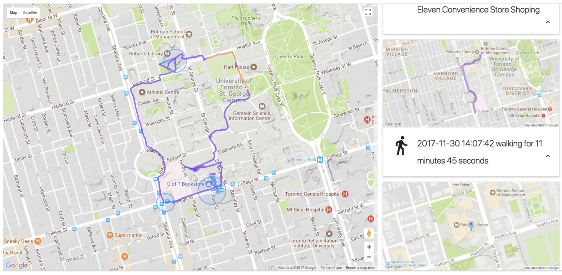


Figure 4. Overall View
We collect the data on Nov 30, 2017, around the St. Geoge Campus of the University of Toronto. This data was collected in one hour, and it represents regular daily schedule of a UofT student.

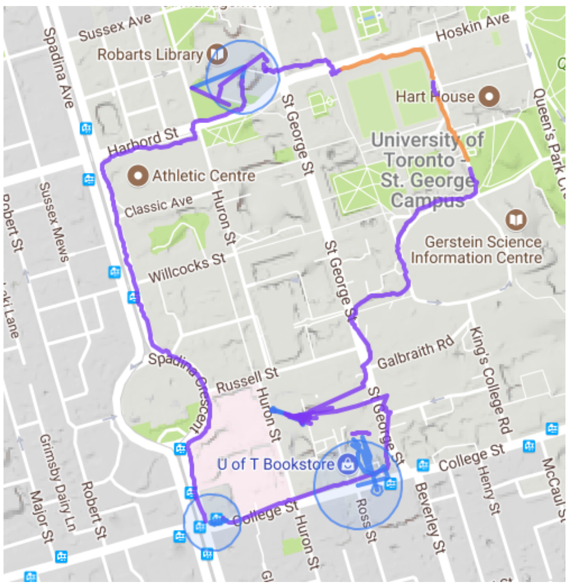


Figure 5. Map View

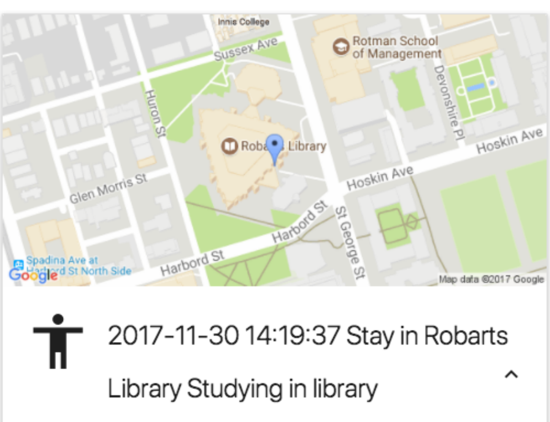


Figure 6. Card

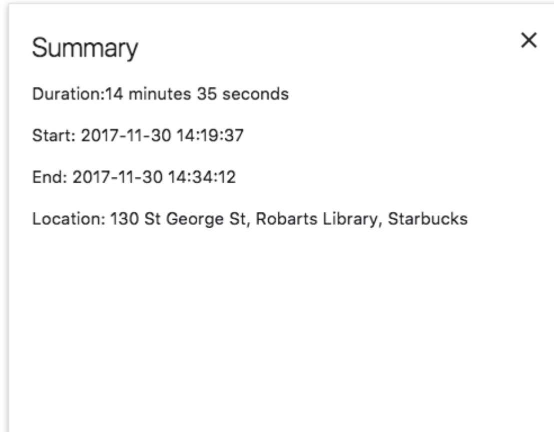


Figure 7. Summary

3. EVALUATION

In this section we present experiments for our activity recognition techniques. Two questions arise for classification, that is how many successive information should we use and what classification algorithms works best? We perform experiments to exploit on the two questions. Evaluation results show that with only tens of successive timestamps' data, our algorithms give activity recognition results with high accuracy. What's more, we also plot the confusion matrices between different activities, for different classification algorithms, which reveals similarity difference between activities.

3.1. Accuracy

From all those existing algorithms, we choose the most popular ones, Logistic regression (LR), k-th nearest neighborhood (KNN), decision trees (DT) and deep neural networks (MLP). For all the experiments, we use k as 5 in KNN. We tune the best performing neural network from one hidden layer of 10, 20, 50 units. We experiment each algorithm on different length N of data. For length N, we concatenate 3-dim data from N successive timestamps, forming a 3N dimensional vector for classification. As the data collection frequency is 50Hz, this vector ranges $\frac{N}{50}$ seconds. For every chosen algorithm, we train the model on training dataset and evaluating on the testing dataset, for N in 1, 3, 5, 10, 20, 40, 50, 70, 90, 100, 120, 130 and 150. Results are shown in Fig 8.

As shown from Fig 8, for all the algorithms, the evaluation accuracy will increase first and then decrease with increasing of N. This is due to the trade-off between feature size and data size. When using bigger N, one feature carries more information, but number of training instances decrease. When N is big enough, the training data size is not big enough to train the model well, therefore the evaluation performance

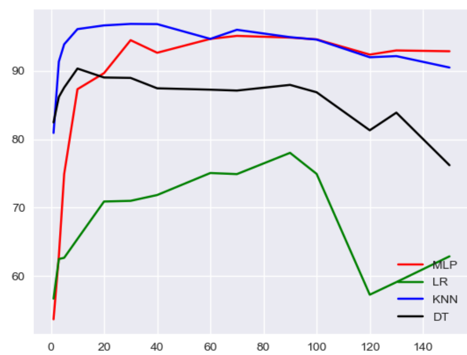
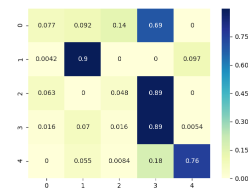
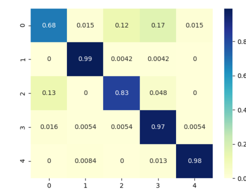


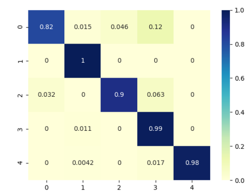
Figure 8. Accuracy



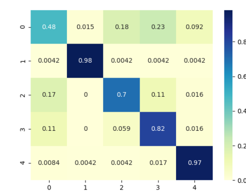
(a) LR



(b) MLP



(c) KNN



(d) DT

Figure 9. Confusion matrix

will drop. Comparing between all these different algorithms, KNN and MLP perform consistently better than decision tree and logistic regression. Although KNN also works better than MLP, it needs to use all the training data, thus can bring along with big memory cost and computation cost for testing. Therefore MLP seems a better choice when the accuracy requirement is not very strong.

3.2. Confusion Matrix

Beyond investigating overall classification accuracies, we also plot the confusion matrix for all these algorithms in Fig 9, which can reveal the relations between different activities. This experiment is performed with N=30. In the figure, 0, 1, 2, 3, 4 correspond to going down stairs, staying still, going up stairs, walking, and running, respectively. And (i, j) block in the figure represents the percentage of class i being classified as class j. As shown from the results, MLP, KNN and DT can classify all these activities with at least moderately high accuracies. While for the logistic regression, large amount of going up stairs and going down stairs are classified as walking. This is due to that logistic regression can only distinguish data instances distributed besides two sides of one hyperplane, while other algorithms can use nonlinearity to attain more complex decision boundary. This result also shows that going up or down stairs are more similar with walking than other activities, which coincides with our general knowledge.

4. RELATED WORK

Activity recognition has long been a eye-catching research area, that many different algorithms and datasets have been proposed to promote the development of activity recognition, towards higher accuracy and more diverse genres. The first works on human activity recognition (HAR) was done in late 90s (Lara & Labrador, 2013). Majority of existing work on HAR used acceleration data from the wearable devices to recognize activities (Lara & Labrador, 2013; Kwapisz et al., 2011). The existing solutions focus on ambulation activities (e.g. Walking, running, sitting), but there are some works that focus on transportation (e.g. Riding a bus, cycling, and driving) and exercise (e.g. Rowing and lifting weights) (Lara & Labrador, 2013). Type of Machine learning model often used for activity recognition are Decision tree, Bayesian, KNN, SVM and neural network (Lara & Labrador, 2013; Yang et al., 2015). Given the more advancement happen recently, there are still challenges exists in the HAR field. For example, the system needs to meet the accuracy and reliability requirements while maintain low energy consumption (Lara & Labrador, 2013).

(Anguita et al., 2012b) presents a hardware-friendly technique by applying Support Vector Machine (SVM) to classify activities. More recently, (Martin et al., 2017b) combines dimension reduction and classification algorithms to achieve both advantages of low energy cost and high recognition accuracy. In parallel, many datasets are published to provide better platform for activity recognition including UniMib (Micucci et al., 2017), MobiAct (Vavoulas et al., 2016) and RealWorld (Sztyler & Stuckenschmidt, 2016). These datasets not only provides data for different activities like walking and running, but also includes data instances for different falling status, which makes it applicable for better falling detection.

Another approach of activity recognition is based on GPS or combination of accelerometer and GPS (Liao et al., 1970; Zheng et al., 2008; 2010; Ermes et al., 2008). (Liao et al., 1970) presents the location based activity recognition and significant place detection using relational Markov network and achieved round 85% of accuracy. Low level of activity recognition is based the detection of the high level significant place. This approach is similar to us but we focus on the what the user is doing instead of low level activity like walking that we recognize using acceleration data. (Zheng et al., 2008) focus on transportation mode dection. Instead of using simple featuer like speed, (Zheng et al., 2008) identifies heading change rate, velocity change rate and stop rate. (Zheng et al., 2010) focus on recommendation system for nearby location based the previsou acivitiy and visited places. (Ermes et al., 2008) is most similar to HARU with focus on dectection of daily sports. It combines accelerometers data and GPS data to detect various kind of indoor and outdoor sports.

5. CONCLUSION

Researchers have been focusing on the use of mobile sensors for a long time.In this work, we present HARU, which performs accurate activity classification and schedule recognition based on continuous collected accelerator data,aiming to explore the possibilities of human activity recognition usage. HARU executes with interaction of mobile phones and cloud. HARU uses self-developed mobile application to collect real-time data and offload to the cloud for efficient.We found that human activities can be recognized with fairly high accuracy using a single accelerometer. The accelerometer data were acquired from multiple subjects under real-world conditions for two most common phone positions: smartphone in hand and smartphone in pants pocket. Different classifiers were used for evaluating recognition performance. This accurate activity recognition provides the possibility for researchers to further use these activity recognition results in the other fields of research. We combine the activity recognition results with location information, which demonstrates what place the user is visiting, to realize the function of record the user's daily schedule and predict the users habit.

Although there are still many limitations of our project, such as the high power consumption, location information positioning bias and unloading data loss with interrupted network, we believe this is still an interesting and meaningful attempt in the usage of human activity recognition. In addition, this project give us an opportunity on applying mobile and cloud computing knowledge in the process of implement this system.

6. FUTURE WORK

For future works, we would like to add more actions capabilities to HARU. Base on the GPS location data we collected. We could calculate distanced travel and speed of travel. Adding this information to the action classifier, we can classify if the user is driving or riding a bike. We could use the method presents by (Liao et al., 1970). Based on the information we collected from the user with some additional information like gender, age and weight we could add more functionalities to HARU. For example, we could estimate the calories consumption of the user. With the foundation we built, we believe there is numerous possibility for HARU. Beside the new features, we could add to HARU; there are some improvements we can make. GPS sensor on iPhone is not accurate. The recorded longitude and latitude can be way off from the true location. We could use the accelerometer data collected and previous GPS data to detect the random noise and remove them. Currently, we only predict what user is doing when the classified action is staying still. We could also predict what is heading to if the user is walking or running. However, this required historical data of the user and more training data for a complex model. More future directions in human activity recognition like predicting future activities and overlapping activities are discussed in the paper(Lara & Labrador, 2013).

Beyond functionality increment, moving computation from cloud to mobile phone can also be useful to explore as it prevents the failure when network is not accessible. But an important factor lies is energy cost, which lives as the core of mobile devices. As known that continuous sensing consumes large amount of energy, improvements can be made by some techniques. For example, with real-time activity recognition, if the user is classified as staying still or going up/down stairs, the geographical location won't change then. Thus we can turn off the energy-consuming GPS monitoring based on the classification results. What's more, different classification algorithms can be combined to reduce energy cost further. As shown in Fig 9, although logistic regression shows low accuracy overall, it achieves high accuracy when the activity is recognized as staying still or running. Therefore, we can use energy-efficient logistic regression as the first filter, only if it outputs classes other than staying still and running, deep neural network is put into use. In practice, this can reduce large amount of energy cost as the two classes take big fraction of user's activity.

References

Anguita, Davide, Ghio, Alessandro, Oneto, Luca, Parra, Xavier, and Reyes-Ortiz, Jorge L. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. *Lecture Notes in Computer Science Ambient Assisted Living and Home Care*, pp. 216223, 2012a. doi: 10.1007/978-3-642-35395-6_30.

Anguita, Davide, Ghio, Alessandro, Oneto, Luca, Parra, Xavier, and Reyes-Ortiz, Jorge L. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *International workshop on ambient assisted living*, pp. 216–223. Springer, 2012b.

Banos, Oresti, Galvez, Juan-Manuel, Damas, Miguel, Pomares, Hector, and Rojas, Ignacio. Window size impact in human activity recognition. *Sensors*, 14(4):64746499, Sep 2014. doi: 10.3390/s140406474.

Bayat, Akram, Pomplun, Marc, and Tran, Duc A. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, 34:450457, 2014. doi: 10.1016/j.procs.2014.07.009.

Ermes, M., Parkka, J., Mantyjarvi, J., and Korhonen, I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Transactions on Information Technology in Biomedicine*, 12(1):2026, 2008. doi: 10.1109/titb.2007.899496.

Kwapisz, Jennifer R., Weiss, Gary M., and Moore, Samuel A. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 12(2):74, 2011. doi: 10.1145/1964897.1964918.

Lara, Oscar D and Labrador, Miguel A. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials*, 15(3):1192–1209, 2013.

Lee, Heyoung, Ahn, Heejune, Choi, Samwook, and Choi, Wanbok. The sams: Smartphone addiction management system and verification. *Journal of medical systems*, 38 (1):1, 2014.

Liao, Lin, Fox, Dieter, and Kautz, Henry. Location-based activity recognition, Jan 1970. URL https://iths.pure.elsevier.com/en/publications/location-based-activity-recognition-2.

Martin, Bryan, Addona, Vittorio, Wolfson, Julian, Adomavicius, Gediminas, and Fan, Yingling. Methods for real-time prediction of the mode of travel using smartphone-based gps and accelerometer data. *Sensors*, 17(9):2058, Aug 2017a. doi: 10.3390/s17092058.

Martin, Bryan D, Addona, Vittorio, Wolfson, Julian, Adomavicius, Gediminas, and Fan, Yingling. Methods for real-time prediction of the mode of travel using smartphone-based gps and accelerometer data. *Sensors*, 17(9):2058, 2017b.

Micucci, Daniela, Mobilio, Marco, and Napoletano, Paolo. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, 7(10):1101, 2017.

Sztyler, Timo and Stuckenschmidt, Heiner. On-body localization of wearable devices: An investigation of position-aware activity recognition. In *Pervasive Computing and Communications (PerCom), 2016 IEEE International Conference on*, pp. 1–9. IEEE, 2016.

Vavoulas, George, Chatzaki, Charikleia, Malliotakis, Thodoris, Pediaditis, Matthew, and Tsiknakis, Manolis. The mobiact dataset: Recognition of activities of daily living using smartphones. In *ICT4AgeingWell*, pp. 143–151, 2016.

Yang, Jianbo, Nguyen, Minh Nhut, San, Phyo Phyo, Li, Xiaoli, and Krishnaswamy, Shonali. Deep convolutional neural networks on multichannel time series for human activity recognition. In *IJCAI*, pp. 3995–4001, 2015.

Zheng, Vincent W., Zheng, Yu, Xie, Xing, and Yang, Qiang. Collaborative location and activity recommendations with gps history data. *Proceedings of the 19th international conference on World wide web - WWW 10*, 2010. doi: 10.1145/1772690.1772795.

Zheng, Yu, Li, Quannan, Chen, Yukun, Xie, Xing, and Ma, Wei-Ying. Understanding mobility based on gps data. *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp 08*, 2008. doi: 10.1145/1409635.1409677.