An Activity-centered Comparison on Human Activity Recognition

Oleg Geier

Seminar KI: gestern, heute, morgen
Angewandte Informatik, Universität Bamberg

Abstract. Human Activity Recognition (HAR) research is focused on increasing the accuracy and precision for a predefined activity set. Adjustable parameters for improvement are the sensor selection, sensor placement and choice of classification algorithm. The HAR task is to determine the corresponding activity for a given time interval on the raw sensor data. This paper covers a comparison based on the selected single activities instead of the complete activity set. The separation is used for an one versus all activity evaluation within the set. We conclude that it would be more helpful to consider the influence between individual activities to improve recognition performance.

Keywords: activity recognition, machine learning

1 Motivation

Human Activity Recognition (HAR) is the task of identifying and correctly classifying an individual activity like walking or cycling from a predefined set of activities. The recognition task is typically done with wearable sensors (smartphones, smartwatches, fitness band, etc.), image-based activity detection (cameras, moving patterns) or external sensors (light barrier, door contact, smartcards, etc.). The literature (Lara & Labrador, 2013) names a list of high variety target sets for possible tracking tasks: ambulation, transportation, phone usage, daily activities, exercise / fitness, military and upper body (see table 1).

Another area for Human Activity Recognition (HAR) is the medical domain. Medical conditions often require supervision. Either by following a detailed plan (taking insulin shots as a diabetic, following a strict diet; obesity / anorexia) or doing regular exercise (dementia, diabetes, obesity). In other cases an early warning is favourable to prevent critical situations that may be triggered by bad air quality or high humidity for example (Centers for Disease Control and Prevention).

In all cases the goal is to liberate patients from stationary hospital care. The patient should be able to monitor his or her situation without the need for long periods of hospitalization. Furthermore doctors can monitor and advise their patients remotely if necessary. The data gathered from these devices can be used to adjust a proper treatment for each patient individually. It could even be used as a guidance for similar cases based on the data from previous patients.

Table 1. Types of activities recognized by state-of-the-art HAR systems. (Lara & Labrador, 2013)

Group	**Activities**
Ambulation	Walking, running, sitting, standing still, lying, climbing stairs, descending stairs, riding escalator, and riding elevator.
Transportation	Riding a bus, cycling, and driving.
Phone usage	Text messaging, making a call.
Daily activities	Eating, drinking, working at the PC, watching TV, reading, brushing teeth, stretching, scrubbing, and vacuuming.
Exercise/fitness	Rowing, lifting weights, spinning, Nordic walking, and doing push ups.
Military	Crawling, kneeling, situation assessment, and opening a door.
Upper body	Chewing, speaking, swallowing, sighing, and moving the head.

Finally devices like smartphones play a crucial role in our everyday life and accompany us in all situations. The medical findings can be incorporated into a smart mobile assistant device. A virtual coach / trainer can give personalized guidance to a healthy life. The need for physical and psychological activity is for every person different as are the capabilities of each person. An always-on, always-there device can learn this individual needs.

One big problem in HAR is the wide variety in experiment design. The used sensors, used machine learning algorithms and used features differ for each set of activities. The need for an unified approach to activity recognition arises quickly. At least identifying the components for the same basic activities shouldn't have to be evaluated over and over again. The aim of this paper is to provide a comparison of activities instead of the underlying activity set.

The following section will cover the basic terminology and the definition of the classification task. In section 3 activities are grouped in similar movement and evaluated based on classification error and accuracy. The last section closes with a conclusion of the results.

2 Background

In a survey on Human Activity Recognition, Lara and Labrador (2013) identified the common problems of HAR. This paper will cover a detailed comparison between *individual* activities and their accuracy at being identified as such in a larger set. A part that is not covered in the survey mentioned above. The general aspects of HAR are covered in this section. Before an activity recognition process can be compared to others it is necessary to lay down the foundation on how activities are identified in general. The definitions are taken directly from Lara and Labrador (2013).

Definition 1. *HAR problem (HARP): Given a set $S = \{S_0, \ldots, S_{k-1}\}$ of k time series, each one from a particular measured attribute, and all defined within time interval $I = [t_\alpha, t_\omega]$, the goal is to find a temporal partition $\langle I_0, \ldots, I_{r-1} \rangle$ of I, based on the data in S, and a set of labels representing the activity performed during each interval I_j (e.g., sitting, walking, etc.). This implies that time intervals I_j are consecutive, non-empty, non-overlapping, and such that $\bigcup_{j=0}^{r-1} I_j = I$.*

Definition 2. *Relaxed HAR problem: Given (1) a set $W = \{W_0, \ldots, W_{m-1}\}$ of m equally sized time windows, totally or partially labeled, and such that each W_i contains a set of time series $S_i = \{S_{i,0}, \ldots, S_{i,k-1}\}$ from each of the k measured attributes, and (2) a set $A = \{a_0, \ldots, a_{n-1}\}$ of activity labels, the goal is to find a mapping function $f : S_i \rightarrow A$ that can be evaluated for all possible values of S_i, such that $f(S_i)$ is as similar as possible to the actual activity performed during W_i.*

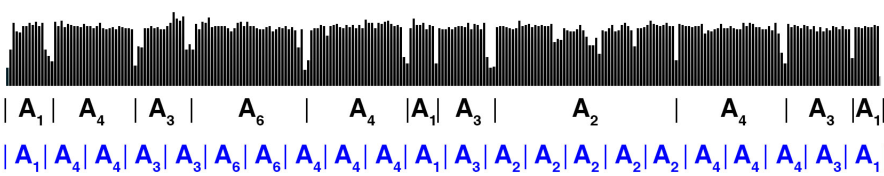

Fig. 1. Partitioning of raw sensor signal into separate activities A_1 to A_6 with Def. 1, HAR problem (black) and fixed window size, relaxed HAR problem (blue).

Both definitions imply that only one activity can be performed during any time interval or time window. Simple activities like walking cannot be combined with other activities like laughing. Whereas in the real world such a joined distinction would be preferable. Figure 1 depicts the classification visually. The time windows of the relaxed problem (blue color) are always the same size whereas with HARP (definition 1) the time points of the transition always vary and have to be found beforehand (black color). Even though the later one would be preferable it is not feasible to compute these transition points (Lara & Labrador, 2013).

On the other hand the relaxed problem (definition 2) reduces the computation complexity but generates a different kind of problem. A time window can now contain multiple activities and have to decide for a specific one. In this case the activity with the longest duration was chosen for each window which results in activities A_4, A_3 and A_6 being the same size after classification (fig. 1). Whereas the second activity A_4 now seems to be larger than A_6. Therefore the transition window from one activity to another is prone to error but neglectable because the number of transition windows is expected to be small, compared to the number of normal activity windows (Lara & Labrador, 2013). The second A_1-activity was detected just by chance. If all windows were shifted by half a window size this segment could be classified as A_4 and A_3 instead; overlooking A_1 completely.

3 Comparison of Trackable Activities

3.1 (Nordic-)Walking, Ascending- and Descending Stairs

One of the most researched activities is without doubt the **walking** activity (Lara, Pérez, Labrador, & Posada, 2012). Most commonly recognized together with lying, standing, sitting, running, as well as ascending and descending stairs. The number of accelerometer sensors vary between one (Hanai, Nishimura, & Kuroda, 2009), two (Riboni & Bettini, 2011) and five sensors (Bao & Intille, 2004). However Bao and Intille concluded that two sensors are sufficient for identifying basic ambulation activities if they are placed at a wrist and either thigh or hip. Achieving accuracy rates between 80% and 95% for 20 everyday activities.

The most obtrusive setup (Pärkkä et al., 2006) with 22 sensors placed throughout the body achieves an accuracy of 99% on the walking activity. They differentiate between lying, running, rowing, walking, nordic walking, sitting/standing and cycling. As with most tracking tasks the problem lies in identifying and distinguishing similar activities like walking and nordic walking. 42.4% of the walking activities are wrongly classified as mostly either nordic walking or sitting/standing (see table 2). It is not intuitively understood how walking can be confused with sitting. The authors say it can be explained with inaccuracies in annotation. For example activity annotated as walking often includes short periods of standing.

Table 2. Confusion matrix of custom decision tree (Pärkkä et al., 2006)

	Lie	Row	Bike	Sit	Run	Nordic	Walk
Lie	1417	0	0	205	0	0	0
Row	0	1646	0	717	0	0	23
Bike	0	0	2461	612	0	0	29
Sit	121	40	53	34083	4	340	962
Run	0	0	0	44	2284	21	5
Nordic	0	1	0	256	39	4507	194
Walk	0	16	4	5412	15	3964	12797

The same problem arises by distinguishing between ascending / descending stairs and walking. Once again, since the activity is too similar to the others, many classification errors are conducted. This similarity in the data is visualized in Randell and Muller (2000) where the data for the selected features is clearly not linearly separable (depicted in figure 2). The sample points overlap each other. The authors still get a high accuracy between 85% and 90% showing that artificial networks achieve similar results to decision trees (Pärkkä et al., 2006) and HAAR-like filters (Hanai et al., 2009). The later achieving up to 93.9% with just one single, chest mounted sensor, also retaining a low calculation cost.

A proper selection of features is therefore absolutely essential. Lara et al. (2012) propose the trend and magnitude of change as two additional transient features to minimize classification error for transition windows. Furthermore windows are no longer consecutive but overlap 50% with the previous and following time window. They achieve an accuracy of 92.9% with Additive Logistic Regression and a time window size of 5 seconds. For 12 second time windows it even gets up to an accuracy of 94.3%.

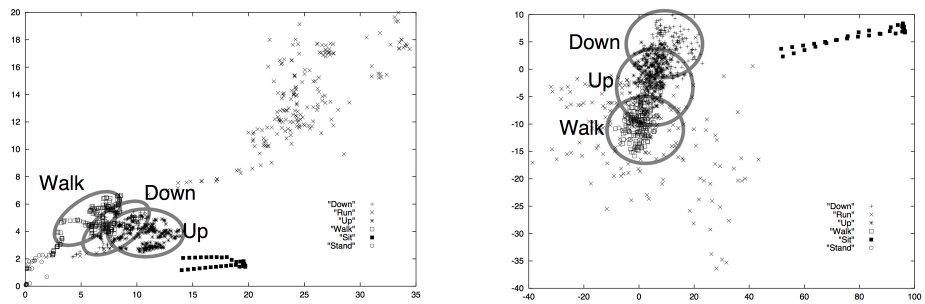

Fig. 2. Similar sensor movement for features RMS (left) and Integrated Acc (right) (Randell & Muller, 2000)

Table 3. Confusion matrix of common activities (Hanai et al., 2009)

	Walking	Running	Standing	Ascending	Descending
Walking	848	0	7	39	14
Running	0	909	0	0	0
Standing	6	0	814	1	1
Ascending	26	0	0	503	59
Descending	18	1	0	58	471

3.2 Running and Jogging

The **running** activity is different to previously mentioned rather calm movements in the sense that sensors are shaken more rapidly. The sensor data is therefore more easy to distinguish from slower walking, sitting or lying activities. The whole body moves more or less synchronous up and down for each running step. There should be a high similarity in sensor signals for all mounted accelerometer sensors. The accuracy achieved by Lara et al. (2012) reflects this with 98.6% and 100% for 5 seconds and 12 seconds respectively.

Riboni and Bettini (2011) compare statistical classification algorithms like Bayesian approaches, decision trees, probabilistic discriminative models and kernel machines against their COSAR approach. With a significant boost in precision for jogging activity. A jump from 71.9% to 96.8%. Just by taking into account the current GPS position and ruling out activities that aren't possible at that location. For example the system won't classify the brushing teeth activity if the user is currently in the meadow or in the woods. This bias is prevalent in the underlying model. The two most often confused activities for jogging in this paper are the standing activity and the walking activity. From 290 and 304 misclassified instances respectively it drops down to 9 instances with COSAR.

To reduce the transition error Khan, Lee, Lee, and Kim (2010) propose eight dedicated transition activities: Lie-Stand, Stand-Lie, Lie-Sit, Sit-Lie, Sit-Stand, Stand-Sit, Walk-Stand, Stand-Stand. Each activity change has to go through one of the predefined and recorded transition activities. The data origin is a single triaxial accelerometer sensor. With the features autoregressive coefficients, signal-magnitude area and tilt angle they achieve an average accuracy of 97.9% overcoming the difficulties others have encountered in these transition phases.

A similar approach is followed in Zhu and Sheng (2009). To extract transition states the authors apply a coarse-grained classification with feature extraction and a neural network for each sensor. The next layer consists of a sensor fusion step combining the results of all neural networks. Thereof transitional and strong displacement activities are detected and passed to the fine-grained classification step. Here heuristic discrimination is used for transition activities and Hidden Markov Models for the actual activities. They achieve an average overall accuracy of 89.7% for the four difficult cases walking, ascending-, descending stairs and running.

3.3 Cycling

Due to the perpetuate rotating movement during **bicycling** one sensor – worn on the ankle – is sufficient for tracking that activity. Ermes, Parkka, and Cluitmans (2008) get an accuracy of 96% for the cycle activity versus the lie, stand, walk or run activity (see table 4). However the cycling was done on an exercise bike and is therefore not necessarily comparable to a real world bicycling experience outside the lab. The authors explain the misclassification to walking and standing with short lasting breaks within the walking periods.

Taking into account a real world scenario the recognition rate can drop significantly. Bao and Intille (2004) conduct an accuracy of 95.8% under laboratory conditions but only 66.7% under a naturalistic setting. Both times using four sensors on chest, thigh, wrist and forearm for 24 subjects. A sensor placed at the foot or ankle is missing completely.

Pärkkä et al. (2006) describe the cycling activity is detected through a left-right movement of the chest. This is also the reason why the activity is confused with the sitting / standing activity. Both features are overlapping slightly resulting in a misclassification.

Table 4. Confusion matrix of the activity classification. The values shown are percentages. (Ermes et al., 2008)

	Lie	Sit/Stand	Walk	Run	Cycle
Lie	100	0	0	0	0
Sit/Stand	0	100	0	0	0
Walk	0	2	96	0	2
Run	0	4	21	75	0
Cycle	0	4	0	0	96

3.4 Jumping

A rather simple activity is described in He, Liu, Jin, Zhen, and Huang (2008). The **jumping** activity is detected as an absence of any accelerometer signal. The authors describe a weightlessness feature where during a short period of time not even gravitational pull is registered due to the fall of the subject. Preceded and succeeded by an increased signal amplitude. With this weightlessness feature they achieve an accuracy of 96.9% compared to traditional time-domain features with an average accuracy of 61.5%.

In later experiments (He & Jin, 2009) they replaced this weightlessness feature with a general Discrete Cosine Transform (Ahmed, Natarajan, & Rao, 1974) and got an even higher accuracy of 97.2% (96.3% for jumping activity exclusively).

3.5 Other activities

Cheng, Amft, and Lukowicz (2010) use four electrodes paced at the chest, wrist, leg and neck to track user activities. The sensor placement at the neck is unusual compared to other studies but allows for additional activity tracking. Namely chewing (79.6% accuracy), **swallowing** and head movement. They can even differentiate between swallowing water (74.9% accuracy) and swallowing bread (90.6% accuracy) because they result in a different signal pattern. The misclassification for swallowing bread when the correct classification should be swallowing water is at 14.5%. The overall upper body movement accuracy is at 77%. The authors also note that including the **walking** activity reduces the performance of the overall system.

Identifying **soldier activities** is discussed in Minnen, Westeyn, Ashbrook, Presti, and Starner (2007). Additionally to the previously mentioned basic activities, these military specific activities are also considered: crawling, kneeling, driving, a weapon up state, shaking hands, opening a door and situation assessment. They didn't go into detail with the individual activities other than the overall accuracy of 90.3%. But another very important topic emerged from the experimental results is manual vs. automatic feature selection. For the shaking hands activity the selected features were apparently taken off the accelerometer sensor from the right wrist as expected. The feature for the driving activity is not intuitively understood. Here the primary feature was taken from high frequencies in the chest and thigh sensor. The authors believe this is due to the vibration of the humvee engine.

Writing and **typing** activities are detected in Pham and Abdelzaher (2008). The complementary use of energy features and relative error energy features results in an accuracy of 96.3% for typing and 93.9% for writing for multiple subjects. Hidden Markov Models reach only 67.7% and 78.5% respectively.

4 Conclusion

Human Activity Recognition is a complex topic with many adjustable subtopics to be answered beforehand. Not only does the selected classification algorithm influence the detection performance but also the selection of the sensors and where to place them plays a crucial role for success. Also the problem of what features have to be extracted from the raw signal is not solved yet and different for each activity. For image data we have sophisticated algorithms extracting features like edges and corners, for audio data the frequency transformation is a predominant feature extraction method. A similar general approach is missing for time based accelerometer sensor data.

The third most important design factor is the selection of activities. We see a huge potential for further research in this area. Most research conducted so far uses a set of everyday activities or a case-dependent set of activities (e.g. soldier movement with military background). Activities should rather be selected by mutual influence. An activity set could instead be constructed such that the classification error is reduced. Adding only activities that are very distinct from others in the set. Is it really appropriate to have very basic activities like standing and somewhat more complex activities like cycling in the same activity set? Generally speaking activities that happen often during other activities like walking or standing should be avoided if high accuracy is preferred.

Human Activity Recognition works great for activities that are very distinct from one another regarding the extracted features of the sensorial output. Whenever activities are very similar like walking, ascending-, descending stairs and nordic walking they will be confused with the others. However good sensor placement and selection of proper feature selection can diminish these misclassifications in some cases up to 35.4%. A loss in accuracy can also happen for activities with distinct body movement. This is most often due to labeling errors made during activity recording. This correlates with the previously mentioned combination of simple and complex events. A sitting activity can happen during cycling if the subject is just rolling downhill. A standing activity can happen during a run if the user makes a short break. The data is later falsely labeled as running even though standing would be the appropriate label.

References

Ahmed, N., Natarajan, T., & Rao, K. R. (1974). Discrete Cosine Transform. *IEEE Transactions on Computers*, *100*(1), 90–93.

Bao, L., & Intille, S. S. (2004). Activity Recognition from User-Annotated Acceleration Data. In *International Conference on Pervasive Computing* (pp. 1–17).

Centers for Disease Control and Prevention. (n.d.-a). *Common Asthma Triggers.* Retrieved Feb 2, 2018, from `https://www.cdc.gov/asthma/triggers.html`

Centers for Disease Control and Prevention. (n.d.-b). *Managing Epilepsy.* Retrieved Feb 2, 2018, from `https://www.cdc.gov/epilepsy/managing-epilepsy/index.htm`

Centers for Disease Control and Prevention. (n.d.-c). *Preventing Epilepsy.* Retrieved Feb 2, 2018, from `https://www.cdc.gov/epilepsy/preventing-epilepsy.htm`

Cheng, J., Amft, O., & Lukowicz, P. (2010). Active Capacitive Sensing: Exploring a New Wearable Sensing Modality for Activity Recognition. In *International Conference on Pervasive Computing* (pp. 319–336).

Ermes, M., Parkka, J., & Cluitmans, L. (2008). Advancing from Offline to Online Activity Recognition with Wearable Sensors. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE* (pp. 4451–4454).

Hanai, Y., Nishimura, J., & Kuroda, T. (2009). Haar-Like Filtering for Human Activity Recognition Using 3D Accelerometer. In *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009. IEEE 13th* (pp. 675–678).

He, Z., & Jin, L. (2009). Activity Recognition from acceleration data Based on Discrete Consine Transform and SVM. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on* (pp. 5041–5044).

He, Z., Liu, Z., Jin, L., Zhen, L.-X., & Huang, J.-C. (2008). Weightlessness Feature — A Novel Feature for Single Tri-axial Accelerometer based Activity Recognition. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (pp. 1–4).

Khan, A. M., Lee, Y.-K., Lee, S. Y., & Kim, T.-S. (2010). A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer. *IEEE Transactions on Information Technology in Biomedicine*, *14*(5), 1166–1172.

Lara, O. D., & Labrador, M. A. (2013). A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys and Tutorials*, *15*(3), 1192–1209.

Lara, O. D., Pérez, A. J., Labrador, M. A., & Posada, J. D. (2012). Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive and Mobile Computing*, *8*(5), 717–729.

Minnen, D., Westeyn, T., Ashbrook, D., Presti, P., & Starner, T. (2007). Recognizing Soldier Activities in the Field. In *4th International Workshop on*

Wearable and Implantable Body Sensor Networks (BSN 2007) (pp. 236–241).

Pärkkä, J., Ermes, M., Korpipää, P., Mäntyjärvi, J., Peltola, J., & Korhonen, I. (2006). Activity Classification Using Realistic Data From Wearable Sensors. *IEEE Transactions on Information Technology in Biomedicine*, *10*(1), 119–128.

Pham, N., & Abdelzaher, T. (2008). Robust Dynamic Human Activity Recognition Based on Relative Energy Allocation. In *International Conference on Distributed Computing in Sensor Systems* (pp. 525–530).

Randell, C., & Muller, H. (2000). Context Awareness by Analysing Accelerometer Data. In *Wearable Computers, The Fourth International Symposium on* (pp. 175–176).

Riboni, D., & Bettini, C. (2011). COSAR: hybrid reasoning for context-aware activity recognition. *Personal and Ubiquitous Computing*, *15*(3), 271–289.

Zhu, C., & Sheng, W. (2009). Human Daily Activity Recognition in Robot-assisted Living Using Multi-sensor Fusion. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on* (pp. 2154–2159).