Recognition of High-Level Activities with a Smartphone

Božidara Cvetković
Violeta Mirchevska
Vito Janko
Mitja Luštrek
Jožef Stefan Institute
Department of Intelligent Systems
1000 Ljubljana, Slovenia
boza.cvetkovic@ijs.si
violeta.mircevska@ijs.si
vito.janko@ijs.si
mitja.lustrek@ijs.si

Abstract
The recognition of high-level activities (such as work, transport and exercise) with a smartphone is a poorly explored topic. This paper presents an approach to such activity recognition that relies on the user's location, physical activity, ambient sound and other features extracted from smartphone sensors. It works in a user-independent fashion, but can also take advantage of activities labeled by the user. It was evaluated on a real-life dataset consisting of ten weeks of recordings. While most activities were recognized quite accurately, the recognition of some revealed two challenges of recognizing diverse lifestyle activities: the ambiguity of some activities, and the inadequacy of smartphone sensors for others.

Author Keywords
High-level activity recognition; lifestyle monitoring; diabetes; smartphone sensors; ECG monitor.

ACM Classification Keywords
[Human-centered computing]: Ubiquitous and mobile computing

Introduction
Knowing the users' activity is useful in a wide range of applications – to understand their context and offer context-sensitive services, to understand their lifestyle and health,

etc. Activity recognition is thus a very active field, which started with dedicated sensors and has lately moved to smartphones and other wearable devices not specifically intended for activity recognition. Our work is motivated by monitoring patients with diabetes, which is a disease strongly linked to the patients' lifestyle. Since many diabetic patients already use a mobile application to manage their disease, using the phone to also recognize their lifestyle activities is a logical extension. It can help the patients themselves to better manage their physical activity and food intake, as well as their physicians to understand the patients' lifestyle.

The activity-recognition approach presented in this paper uses the sensors in the smartphone and an optional ECG monitor (introduced for the management of cardiovascular co-morbidities of diabetes). The key features extracted from the sensors are the user's location, physical activity and ambient sound. These are fed into a general and a user-specific classifier, whose outputs are combined by heuristic rules into the user's final activity. The approach recognizes exercise and eating, which are activities particularly important for diabetic patients, and work, home, out, transport and sleep, which paint a broad picture of the patient's lifestyle and provide context for the ECG and blood-glucose readings.

The following sections of the paper discuss related work, describe our activity-recognition approach, present its experimental evaluation, and conclude the paper.

Related Work

While activity recognition with a smartphone or similar wearable sensors is a mature field, most of the work deals with low-level activities such as sitting, walking and running [5]. The recognition of high-level activities, as in our work, remains largely unexplored, although it provides complementary information to low-level activities. It is of course debatable which activities should be considered low- and which high-level, but our interpretation is that low-level activities are either static (postures) or repeat the same motion pattern on a scale of seconds (e.g., walking).

Low-level activities are typically recognized by passing a sliding window over a stream of acceleration data, extracting a number of features from each window, and feeding the features into a classifier that outputs the activity. Dernbach et al. [6] used the same approach for high-level activities, but they reached the accuracy of barely 50 %, even though they attempted to recognize only a small set of simulated activities. Lee & Cho [9] applied hierarchical hidden Markov models to accelerometer data to first determine low-level activities and from those high-level activities (shopping, taking bus, moving by walk). They reached the precision of around 80 %, but their set of activities was very limited and the users carried the phone in their hand, so the dataset was not a good representation of real life. Garcia-Ceja & Brena [7] recognized commuting, working, at home, shopping and exercising by representing high-level activities with histograms of low-level activities. They reached the accuracy of 80–90 %, but their experiments were user-specific with only one user involved.

With location sensing (GPS, mobile networks) one can determine the users' location, which serves as an important clue to their activity. This way, Lin [10] classified work, sleep, leisure, visit, using a car and other with conditional random fields. He achieved the accuracy of 86 %. Choujaa & Dulay [4] scanned nearby Bluetooth devices to help localize users indoors and recognize additional activities, such as using a computer. They represented the users' daily routine with temporal graphs, requiring manual labeling of each

user's activities. The temporal graphs allowed improving the F-measure of activity recognition by 20 percentage points compared to assuming the same routine as in the training data every day.

Motion and location sensing can be combined with the microphone, visible Wi-Fi networks and light. Wang et al. [13] used such a combination to determine the users' state (working, home_talking, place_speech etc.) with a rule-based system. They achieved the accuracy of around 90 %, but it should be noted that the users' home and office Wi-Fi network names were known apriori to the system, and several of the states were specifically adapted to the system's capability to analyze ambient sounds. Another similar approach was by Miluzzo et al. [12], who tried to infer the users' activity to post about it on a social network.

Our approach does not reach the accuracy of some of the work mentioned above, but it tackles a more difficult problem: (1) it attempts to recognize all the users' activities in real life, including ambiguous ones (e.g., cycling can be exercise, transport or a part of shopping); (2) the activities are not selected to fit the available sensing modalities (unlike some related work, which adapted the activities/states to location or sound sensing); and (3) our approach does not need labeling from each user.

Features for Activity Recognition

- Wi-Fi – location
- GPS – velocity and place category
- Sound
- Acceleration – low-level activity and energy expenditure
- Heart rate
- Respiration

Activity-Recognition Approach

Our activity-recognition approach has two main steps described in the following two subsections: feature computation and machine-learning procedure.

Feature Computation

The features are extracted from typical smartphone sensors and optionally from an accelerometer-equipped chest-worn ECG monitor. They are computed over one-minute windows. An overview of the features is provided in the sidebar, and a detailed description in the rest of this subsection.

Wi-Fi feature has three possible semantic location values computed from the visible Wi-Fi access points: home, work and elsewhere. It is computed in two steps, which require that a part of each user's recordings are used for training (we used one week). In the first step, each phone's Wi-Fi scan in the training dataset is represented by a vector whose values are the signal strengths of the visible Wi-Fi access points. These vectors are clustered, so that each cluster corresponds to a location characterized by visible access points (see our previous work [11] for details). Since the clusters are very user-specific, they cannot be used to build a general activity-recognition classifier. Therefore, the second step transforms them into the three semantic locations which can be used for every user:

- For each Wi-Fi location cluster, the fraction of time the user spends in it during each day is computed.
- Daily important clusters are defined as the five clusters in which the user spends the most time in the day.
- The entire week is divided into working and free days. Working-day importance of each cluster is defined as the number of times the cluster was daily important during working days. Free-day importance is defined analogously.
- Clusters with working-day importance above 0, which are never visited on free days, belong to the location work.
- Clusters with the highest or tied for the highest free-day importance (it can be at most 2, since there are

two free days in a week), which also have working-day importance above 0, belong to the location home.

- All other clusters belong to the location elsewhere.

The Wi-Fi feature corresponds to the correct semantic location with the accuracy of 85 %.

GPS features are again not user-specific: (i) velocity, (ii) category of the nearest place using the Foursquare service API [3] and (iii) whether the user is outdoors or indoors according to the presence or absence of the GPS signal.

Sound features are extracted from the ambient sound recorded with the smartphone's microphone using the jAudio library [1]. We record only 100 ms of sound out of each second to preserve the user's privacy. The recordings are further split into 20 ms sub-windows. The features are the average spectral-centroid, zero-crossing, mel-frequency-cepstral-coefficient (MFCC), linear-predictive-coding (LPC) and method-of-moments values for each sub-window within each one-minute window.

Acceleration features are extracted from the accelerometer (smartphone's and/or ECG monitor's). They are user's most common low-level activity within each one-minute window (low-level activities are computed in two-second windows), and the user's average expended energy (the energy is computed in ten-second windows). They two features are computed with our recent method [5] that can use the smartphone, ECG monitor or both, and can automatically adapt to any orientation and location of the phone on the body. The expended energy is expressed in MET (Metabolic Equivalent of Task, 1 MET corresponds to the energy expended at rest).

Heart-rate features are extracted from the ECG monitor if present. The features are the (i) minimum, (ii) maximum and (iii) average heart-rate within each one-minute window.

Respiration-rate features are also extracted from the ECG monitor. The features are the (i) minimum, (ii) maximum and (iii) average respiration-rate within each one-minute window.

Even though the last three categories of features all use the ECG monitor, the accuracy of the activity recognition is not much degraded without it, so the ECG monitor is not essential for our activity-recognition approach.

Machine-Learning Procedure

The machine-learning activity-recognition procedure utilizes two classifiers: a general classifier trained on data of people other than the user, and an optional user-specific classifier

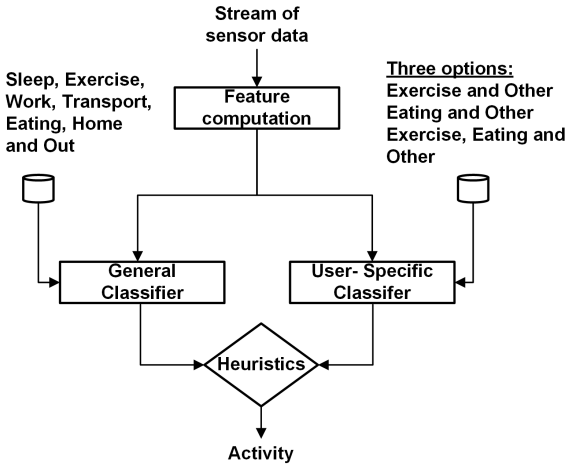

Figure 1: Workflow of the machine-learning procedure.

trained on data labeled by the user. The user does not have to label all the activities: we selected eating and/or exercise, since they are important for diabetic patients and difficult to recognize. If the user chooses to label his/her own data, heuristic rules are used to select the final activity. The general workflow of the procedure is presented in Figure 1.

Algorithm 1 Heuristics rules

```
1:  activities[]                    ▷ Last three recognized activities
2:  general                         ▷ General classifier
3:  userspecific                    ▷ User-specific classifier
4:  MET                             ▷ User's estimated energy expenditure
5:  procedure RecognizeActivity (instance)
6:      aP ← prevalent(activities[])
7:      aG ← general(instance)
8:      aS ← userspecific(instance)
9:      if aG = aP or aS = aP then
10:         return aP
11:     else if aG = exercise or aS = exercise then
12:         if (MET > 2.5) then
13:             return exercise
14:         else
15:             return aG
16:         end if
17:     else if aS = eating then
18:         return eating
19:     else
20:         return aG
21:     end if
22: end procedure
```

Both the general and user-specific classifier are meta-classifiers outputting the majority vote from seven base classifiers trained with machine-learning algorithms implemented in the Weka machine-learning suite [8]: Naive Bayes, Simple Logistic, Support Vector Machine, J48, Random Forest, JRip and AdaBoost. The classifiers were selected empirically.

The heuristic rules used in the machine-learning procedure are presented as Algorithm 1. The recognition is first smoothed by comparing the most common activity among the last three recognized activities against the current outputs of the classifiers. If any of the outputs matches the most common activity, that activity is returned, since in most cases the current activity is the same as the previous one. Otherwise, the next rule checks whether any of the classifiers recognized exercise and the MET value is above 2.5 (the energy expended during leisurely walking). If both criteria are met, exercise is returned. If not, the next rule checks whether the user-specific classifier recognized eating, since eating habits are fairly user-specific. If so, eating is returned; otherwise the output of the general classifier is returned.

Experimental Evaluation

Our activity-recognition approach was evaluated on a dataset of recordings by five volunteers, two weeks by each. The leave-one-person-out method was used for the general classifier: it was trained on the data of four people and tested on the fifth, repeated once for each person. For the user-specific classifier, we used the first week of user-specific data for training and the second week for testing. Since the Wi-Fi feature also required one week for training, all the results are presented for the second week.

Dataset

During the recording, each of the five volunteers (four male, one female) led their life as usual. While some of them had fairly regular daily routines, others had unusual eating patterns, were staying at a different place during the week and

Approach	Activities (F-score)							Average
	Sleep	Exercise	Work	Transport	Eating	Home	Out	
Our previous user-specific approach	0.91	0.49	0.92	0.56	0.29	0.79	0.47	0.63
General	0.91	0.41	0.94	0.66	0.24	0.77	0.58	0.65
General + eating	0.91	0.41	0.94	0.66	0.32	0.77	0.59	0.66
General + exercise	0.92	0.50	0.94	0.72	0.29	0.79	0.62	0.68
General + eating + exercise	0.92	0.50	0.94	0.72	0.34	0.79	0.62	0.69

Table 1: Comparison of different activity-recognition approaches.

weekend, took trips etc., so the resulting dataset is quite challenging from the activity-recognition perspective. The volunteers were asked to carry the smartphone as much as possible, in any pocket they wanted (or in a bag). They were also asked to wear the ECG monitor each day until the battery ran out. On average, we collected 7.5 hours of recordings per day with the ECG monitor and 11 hours with the phone.

We developed a mobile application that recorded all the sensor data needed to compute the features used in our activity-recognition approach. The volunteers also used this application to label the following activities: home-chores, home-leisure, food preparation, eating, exercise, work, out-errands, out-leisure and transport. We later merged home-chores, home-leisure and food preparation into home, and out-errands and out-leisure into out, since these activities proved impossible to distinguish. The volunteers were provided with guidelines regarding labeling, and the application allowed correcting mistakes, but some inconsistencies and mistakes in the dataset certainly remain, and are quite difficult to detect and correct.

Results

We first evaluated our approach using the general classifier only, which does not require the user to label any data. We then added the user-specific classifier recognizing only eating, only exercise and both. All the approaches were compared against our previous work [11], which is completely user-specific: trained on the dataset from the first week and tested on the second week. The results of the comparison are presented in Table 1. We can observe that the general approach outperformed our user-specific approach, probably because it had considerably more training data (from four people vs. from one). The addition of the user-specific classifiers proved beneficial, improving the recognition of the activities it was trained to recognize as well as rhe overall recognition.

The recognition of exercise and eating was still rather poor, even with the user-specific classifier. The confusion matrix in Table 2 shows that exercise is confused with home, out and transport. This is mainly due to the ambiguity of exercise: household chores – if sufficiently strenuous – are in fact exercise, but were not labeled as such because their intent was not exercise; a walk or cycling can be an errand or transport or exercise – this is again a matter of intent. Eating is most often confused with home and work. The confusion with home is perfectly understandable, since sitting at the kitchen table appears very similar when eating, writing, reading or doing any number of other activities. The confusion with work is due to eating at the workplace.

	Recognized						
True	Sleep	Exercise	Work	Transport	Eating	Home	Out
Sleep	**610**	0	0	0	0	68	0
Exercise	0	**277**	3	65	0	142	106
Work	3	8	**6734**	110	74	30	339
Transport	0	15	10	**1030**	3	32	302
Eating	0	0	84	27	**158**	376	172
Home	34	13	11	5	31	**3516**	967
Out	0	108	123	228	7	119	**2019**

Table 2: Confusion matrix of the General + eating + exercise approach.

While recording our dataset, we noticed that the phone battery barely lasted the whole day even though the phone was not used for calls, messaging etc. Since most of the high-level activities we are recognizing typically last several minutes or even hours, we decided to test whether they need to be recognized every minute. We simply recognized the activity in one minute, and then assumed the activity will remain unchanged for the following 5, 10 or 15 minutes. The results for the general classifier only, and for the general + specific classifier recognizing exercise and eating, are shown in Figure 2. The longest activities (sleep, work, home and out) were hardly affected by the sparse recognition. For exercise and transport, the recognition performance decreased with increasing delay between recognitions, which was to be expected, but recognizing the activity every 5 minutes still performed very well. Eating was in some cases even helped by the sparse recognition, because for many meals only a few minutes were recognized correctly, which the sparse recognition smoothed.

Conclusion

In this paper we presented an approach to the recognition of high-level activities with a smartphone, a problem rarely tackled in the activity-recognition field. The approach was tested on a real-life dataset and achieved the F-score of 0.69. It was fairly successful on the activities strongly characterized by location, while it had difficulties with exercise and eating. In the first case, the main reason is that the definition of exercise is subjective. This problem could be tackled by adapting the definition of exercise to the sensors, i.e., by considering every sufficiently vigorous activity to be exercise, although that would not necessarily provide the insight into the user's lifestyle we want. In the second case, the smartphone simply does not have the right sensors to recognize eating. This problem could be solved by a wrist-worn device such as a smart watch, although they are not nearly as ubiquitous as smartphones. In the future we will attempt to improve the recognition of eating with more advanced sound processing, since sound appears to be the only type of data that can be collected with a phone with some chance of doing so.

We also plan to increase the size of our dataset by five more people, and manually clean the data to correct some of the labeling mistakes. Afterwards we will make the dataset available to the community in our repository [2]. We also intend to release the mobile application that was used to col-

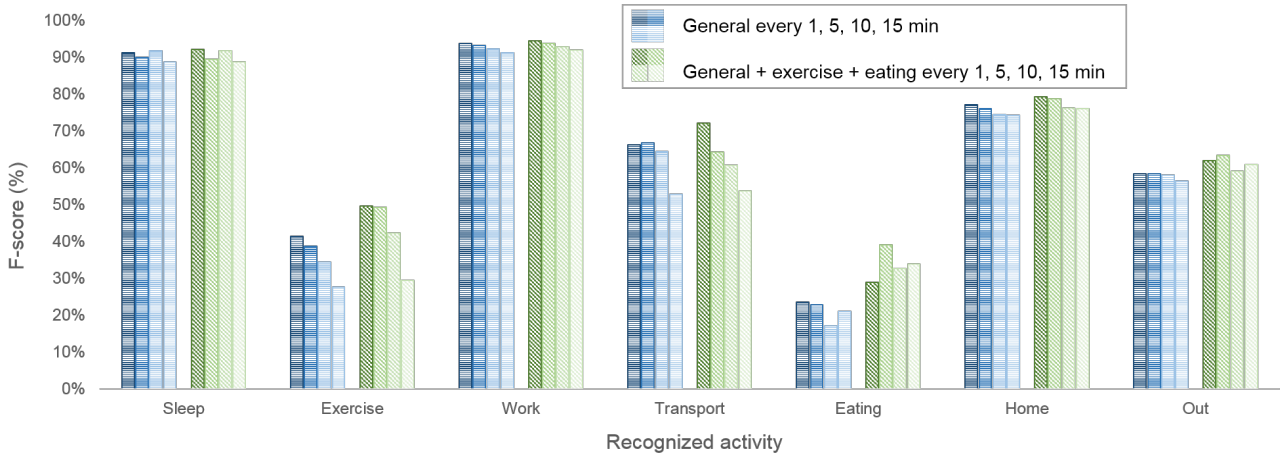

Figure 2: Performance of sparse activity recognition.

lect the data, and that currently recognizes low-level activities and estimates the user's energy expenditure. Whether the recognition of high-level activities will be integrated in this application or will be done on the server is yet to be decided.

Acknowledgments

This work has been partially supported by the EU project COMMODITY12 (`www.commodity12.eu`). We would also like to thank the volunteers who recorded the dataset.

REFERENCES

1. 2011. jAudio library. http://jaudio.sourceforge.net/. (2011).
2. 2013. AmI Repository. http://dis.ijs.si/ami-repository/. (2013).
3. 2014. Foursquare API. https://developer.foursquare.com/. (2014).
4. Driss Choujaa and Naranker Dulay. 2008. TRAcME: Temporal Activity Recognition Using Mobile Phone Data. In *Proceedings of the 2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*. IEEE, New York, USA, 119–126.
5. Bozidara Cvetkovic, Vito Janko, and Mitja Lustrek. 2015. Demo abstract: Activity recognition and human energy expenditure estimation with a smartphone. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*. 193–195. `DOI:` http://dx.doi.org/10.1109/PERCOMW.2015.7134019
6. Stefan Dernbach, Barnan Das, Narayanan C. Krishnan, Brian L. Thomas, and Diane J. Cook. 2012. Simple and Complex Activity Recognition through Smart Phones. In *Proceedings of the Eighth International Conference on Intelligent Environments*. IEEE, New York, USA.

7. Enrique Garcia-Ceja and Ramon Brena. 2013. Long-Term Activity Recognition from Accelerometer Data. In *Proceedings of the 3rd Iberoamerican Conference on Electronics Engineering and Computer Science (Procedia Technology)*, Vol. 7. Elsevier, Amsterdam, The Netherlands.

8. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11, 1 (2009), 10–18. DOI: http://dx.doi.org/10.1145/1656274.1656278

9. Young-Seol Lee and Sung-Bae Cho. 2011. Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer. In *Hybrid Artificial Intelligent Systems*. Springer, Berlin Heidelberg, 460–467.

10. Liao Lin. 2006. *Location-Based Activity Recognition*. Ph.D. Dissertation. University of Washington.

11. Mitja Luštrek, Božidara Cvetković, Violeta Mirchevska, Özgür Kafalı, Alfonso E. Romero, and Kostas Sathis. 2015. Recognising Lifestyle Activities of Diabetic Patients with a Smartphone. In *Proceedings of PHSCD Workshop, Pervasive Health 2015*. ACM, New York, USA.

12. Emiliano Miluzzo, Nicholas D. Lane, Kristóf Fodor, Ronald Peterson, Hong Lu, Mirco Musolesi, Shane B. Eisenman, Xiao Zheng, and Andrew T. Campbell. 2008. Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the CenceMe Application. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*. ACM, New York, USA, 337–350.

13. Yi Wang, Jialiu Lin, Murali Annavaram, Quinn A. Jacobson, Jason Hong, Bhaskar Krishnamachari, and Norman Sadeh. 2009. A Framework of Energy Efficient Mobile Sensing for Automatic User State Recognition. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*. ACM, New York, USA, 179–192.