Sensing Healthy Lifestyle in Urban and Rural Environments

Bert Arnrich
Boğaziçi University
Bebek, Istanbul 34342
Turkey
bert.arnrich@boun.edu.tr

Naime Şeyma Erdem
Boğaziçi University
Bebek, Istanbul 34342
Turkey
seyma.erdem@boun.edu.tr

Hasan Faik Alan
Boğaziçi University
Bebek, Istanbul 34342
Turkey
hasan.alan@boun.edu.tr

Cem Ersoy
Boğaziçi University
Bebek, Istanbul 34342
Turkey
ersoy@boun.edu.tr

ABSTRACT

An extended framework for Android smartphones is employed to obtain objective and subjective measures of relevant lifestyle parameters. Participants were recruited from two very dissimilar environments: the city of Istanbul as one of the largest mega cities worldwide and the rural district of Mersin in the south-eastern part of Turkey. Correlation analysis between subjective ratings and objective sensor data features is performed to illustrate the potential of an automatic detection of health-relevant lifestyle parameters.

1. INTRODUCTION

Major determinants of many health outcomes are manifested in people's lifestyle and behavior [1]. For example, it is widely accepted that indicators of individual's lifestyle and behavior such as amount of physical exercise, amount and quality of sleep, metabolism and socialization patterns are connected to a wide range of health problems such as high-blood pressure, diabetes and mental disorders. It is known that positive health effects can be achieved when these indicators are kept in healthy ranges. An automatic detection of health-relevant behaviors is of particular interest for rural regions where the medical infrastructure is often far less developed in comparison to metropolitan areas and as a result it is tedious to detect health trends in a timely manner.

In this contribution, a feasibility study towards sensing lifestyle parameters in urban and rural environments is presented. Relevant lifestyle parameters are measured and compared between urban and rural environments with the help of participant's own smartphones. An extended framework for Android smartphones is employed to obtain subjective ratings on lifestyle parameters on the one side and to collect objective sensor data on the other side. Correlation analysis between subjective data and sensor features is performed to illustrate the potential of an automatic detection of health-relevant lifestyle parameters.

2. DATA COLLECTION

Data are collected from two groups of subjects who live in two very dissimilar environments: the city of Istanbul as one of the largest mega cities worldwide and the rural district of Mersin in the south-eastern part of Turkey (see also Figure 1). Both groups consist of five people. Average age is 26 years in the urban group and 30 years in the rural group. Data collection is performed during five consecutive days. Three days are weekdays and the remaining two days are weekend days.

Data is collected on the participant's own smartphone. An Android app was developed which allows to collect phone's sensor data in the background and asks for subjective ratings from time to time. The sensor data collection is based on the funf open sensing framework developed by MIT [2]. In addition, a survey mechanism was implemented to obtain subjective ratings. Via Email the link to the Andoid app was sent to the participants. By following this link, the participants installed the app on their own smart phone.

The app stores the data in SQLite databases and sends them to the server from time to time. Data includes device id, timestamp information and a data table which consists of sensor data and survey answers.

Figure 1 Subjects in the first group live in Istanbul and subjects in the second group live in Erdemli located in the rural south-eastern district of Mersin

2.1 Mobile Phone Sensing

In Table 1 all mobile phone sensing modalities that were collected in this contribution are listed. For each modality the amount of data that was collected is shown.

ICCASA 2014, October 15-16, Dubai, United Arab Emirates

DOI 10.4108/icst.iccasa.2014.256949

Acceleration data was recorded to measure the physical activity level of the participants which is known as one of the most important healthy lifestyle parameter. GPS recordings and WiFi fingerprints are obtained for localization purposes. In particular the location data allows to detect important places like home and work. With this information one can calculate working times and how much time is needed for commuting from home to work. Bluetooth is obtained to quantify the number of devices around. This number is known to be a marker of persons around which provides an estimate of social interactions. The screen on/off events are valuable to estimate sleep duration of the user.

Table 1 Amount of data collected for each sensing modality

Sensing Modality	**Data Amount**
Accelerometer	10sec recording every 5min
GPS	One scan every 30 min
WiFi	One scan every 10 min
Bluetooth	One scan every 10 min
Screen On/Off	Event-based

2.2 Mobile Phone Daily Survey

Each participant had to fill out a daily survey consisting of 11 questions related to daily activities and lifestyle. Users are asked to fill out this short questionnaire each evening at 21:00 on their smart phone. In the following we describe the survey in detail.

Physical Activity Level: In order to determine the amount of the physical activity levels walking and running, we asked two questions. The first question was "How long did you walk today?" and the second one was "How long did you run today." For both questions, the participants were asked to select one out of seven choices. In case of running we added an additional choice "None". The seven common choices were as follows:

1. 0 – 20 min
2. 20 – 40 min
3. 40 min – 1 h
4. 1 h – 1 h 20 min
5. 1 h 20 min – 1 h 40 min
6. 1 h 40 min – 2 h
7. More than 2 h

Working Time: "How long did you work today?" was asked for collecting the duration of working time. The participants could select on out of the following seven choices:

1. 0 – 2 h
2. 2 – 4 h
3. 4 – 6 h
4. 6 – 8 h
5. 8 – 10 h
6. 10 – 12 h
7. More than 12 h

Commuting Time: Two questions were asked to gather the commuting time from home to work and back to home: "How long did it take to go to your work place?" and "How long did it take to come back to your home?". For both questions the participants could select one out of 7 choices:

1. 0 – 30 min
2. 30 min – 1 h
3. 1 h – 1 h 30 min
4. 1 h 30 min – 2 h
5. 2 h – 2 h 30 min
6. 2 h 30 min – 3 h
7. More than 3 h

Social Interaction: We asked a total of three questions in order to quantify the amount of three types of social interactions. The first questions was "How many people did you meet today face-to-face?", the second one was "How many people have you communicated today via phone call?", and the third one was "How many people have you communicated today via text messaging?". For each of the three questions the participants were asked to select one out of the following 7 choices:

1. 0 – 3 people
2. 3 – 6 people
3. 6 – 9 people
4. 9 – 12 people
5. 12 – 15 people
6. 15 – 18 people
7. More than 18 people

Sleep Duration: "How many hours did you sleep today?" was asked for gathering sleep duration. The participants had to select one out of the following seven choices:

1. 0 – 2 h
2. 2 – 4 h
3. 4 – 6 h
4. 6 – 8 h
5. 8 – 10 h
6. 10 – 12 h
7. More than 12 h

Stress Level: Following the work on stress detection with mobile phones presented in [4], we asked the participants two questions in order to appraise their experienced stress level. The first question "How stressed have you felt today?" had to be answered by selecting one out of the following five choices:

1. Very low
2. Low
3. Normal
4. High

5. Very High

The second question "How stressed have you felt today compared to yesterday?" had to be answered by selecting on out of the following five choices:

1. Much less: Today, stress level was much lower than yesterday.
2. Lower: Today, stress level was lower than yesterday.
3. Same: Today, stress level was the same like yesterday.
4. Higher: Today, stress level was higher than yesterday.
5. Much more: Today, stress level was much higher than yesterday.

3. RESULTS

In the following we first present results on data completeness. Next, we show the results of the daily survey and if applicable we demonstrate the correlation between survey and sensor data.

3.1 Data Completeness

In Table 2 we present the data completeness for all sensing modalities and the daily survey for both urban and rural groups respectively. For all sensing modalities the data completeness is always lower in the rural group. An explanation for the lower data completeness of the modalities WiFi and Bluetooth can be given by the fact that in the rural part there were less access points available, e.g. the daily average number of nearby WiFi access points in Istanbul was 186 while the daily average was 57 in Erdemli. Each scan which provided no access point was counted as non-complete. The lower rural data completeness rates of the modalities Accelerometer and GPS cannot be explained so far. A potential reason might be the differences in phone models used in the rural group which might result in less complete data. The data completeness of the daily survey is almost identical in the urban and rural groups.

Table 2 Data completeness for all sensing modalities and the daily survey for subjects from urban and rural parts respectively

Sensing Modality	Completeness Urban (%)	Completeness Rural (%)
Accelerometer	87.64	74.48
GPS	71.33	42.36
WiFi	30.47	20.87
Bluetooth	24.50	8.26
Daily Survey	76.0	75.0

3.2 Physical Activity Level

We analyzed the physical activity levels during weekend and weekdays separately for each of the two groups rural and urban respectively. In the left side of Figure 2 it is shown that walking time in the rural environment is higher during weekdays and weekend days. During the weekend, the rural walking time is decreased compared to weekdays. In contrast, in the urban environment the walking time is slightly increased during the weekend. In the right side of Figure 2 we can observe that running time is in general very low for the participants from Istanbul. Participants from Erdemli spent a low amount of time for running during weekdays. Apparently the average amount of walking and running is higher in the rural part of our study. A possible explanation might be the different way of commuting. In Istanbul people usually use their own car or public transportation to reach work, school or home. In Erdemli people usually choose walking for commuting since it is a small district. The decrease in the amount of walking and running during weekend is most probably due to the fact that people usually spend their time on weekends at home with less active tasks.

From the accelerometer data we extracted 14 features such as average values and variance of magnitude, x, y and z axes for each day. The linear correlation coefficients between activity level features and amount of walking from the survey results are presented in the table below. The highest value of 0.74 is found for the feature variance of x-axis acceleration. In Figure 3 the relationship between variance of x-axis acceleration and walking time is shown as a graph.

Table 3. Linear correlation coefficients between activity level features and amount of walking from the survey

Avg Mag	Var Mag	Avg X	Avg Y	Avg Z	Var X	Var Y	Var Z
0.09	0.53	-0.31	0.16	-0.08	0.74	0.49	-0.18

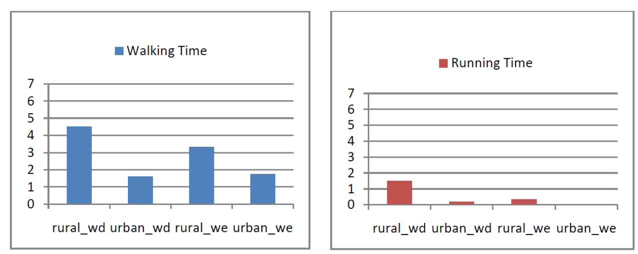

Figure 2 Walking and running level during weekdays (wd) and weekends (we) obtained from the daily survey. Detailed level definition is provided in the section Data Collection: (1)=0 – 20 min, (7)=more than 2 h

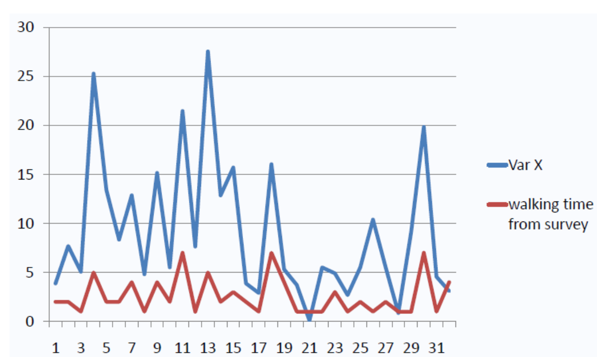

Figure 3 Walking time from the daily survey vs. variance of x-axis acceleration for the corresponding day

3.3 Working Time

In Figure 4, the average values of working time for rural and urban areas on weekdays and weekends are shown. The amount of daily working time in the urban environment is 6-8 hours for weekdays while the average in weekdays is 8-10 hours in rural environment. By looking the difference it can be said that in our study the rural participants worked slightly more than the urban ones. Working duration in both groups decreased during the weekend as expected. However, like during weekdays rural participants worked more on these days and the decline during weekend is less than in the urban part.

In our setup, an assessment of working time can in principle be achieved by detecting the working place of a participant via GPS and/or WiFi fingerprinting. Due to the low amount of available GPS and WiFi data in the rural group, we omitted this kind of assessment.

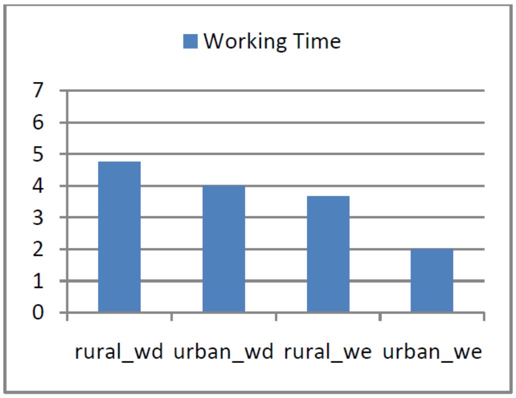

Figure 4 Working time during weekdays (wd) and weekends (we) obtained from the daily survey. Detailed level definition is provided in the section Data Collection: (1)=0 – 2 h, (7)=More than 12 h

3.4 Commuting Time

Since most of our subjects didn't work on weekends, we considered commuting time only for the weekdays. As shown in Figure 5 we observe that subjects from the rural part reach their home or work place within one hour while subjects from the urban part have to travel 1 to 2 hours.

Similar to the working time highlighted above, in our setup an assessment of commuting time can in principle be achieved by detecting home and working place of a participant via GPS and/or WiFi fingerprinting. When relying on GPS one can even assess speed of commuting and thus infer to the traffic situation. Due to the low amount of available GPS and WiFi data in the rural group, we omitted this kind of assessment.

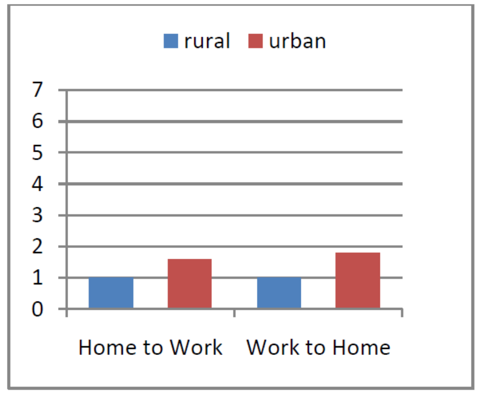

Figure 5 Commuting time obtained from the daily survey. Detailed level definition is provided in the section Data Collection: (1) 0 – 30 min, (7) more than 3 h

3.5 Social Interaction

In Figure 6 is shown that face to face communications is higher in the rural part. During the week, rural people meet more than 15 people per day on average while urban people meet less than 9 people.

Recently, in [3] a face-to-face proximity estimation was introduced which relies on Bluetooth signal strength. In our work, we investigated the number of Bluetooth devices around as a marker of people around. We determined the number of nearby Bluetooth device per day and correlated these numbers with the amount of people that was obtained from the daily survey. The resulting linear correlation coefficient was found to be -0.41. This negative correlation coefficient provides evidence that the number of Bluetooth devices around might be a good estimate of the number of people around but it does not positively correlates with the number of face to face communication. In an urban area one might find many Bluetooth devices around but this does not necessarily mean that one has many face to face communications.

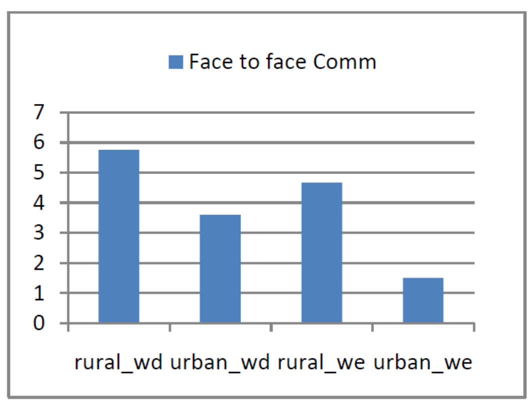

Figure 6 Amount of people involved in a face to face communication during weekdays (wd) and weekends (we) obtained from the daily survey. Detailed level definition is provided in the section Data Collection: (1) 0 – 3 people, (7) more than 18 people

3.6 Sleep Duration

According to our survey results, participants living in rural environment sleep less on weekdays than the people in urban environment. Sleep duration of both groups increases slightly at the weekend as we expected. Average sleep duration is between 4-6 hours in rural and 6-8 hours in urban area during the week.

In order to assess sleep duration automatically we rely on the fact that there should be no phone usage during sleep. We investigated screen on/off events on a daily basis from 10 pm to 11 am of the next day. In case the time interval between two successive screen on-off events is larger than 70 minutes, we assume that the user was falling asleep. With the next screen on-off event we assume that the user is awake. Following this procedure we estimated the sleep duration of each participant and correlated it with the answer from the daily survey. The linear correlation coefficient was found to be 0.11 which is lower than we expected.

3.7 Stress Level

Daily stress level for both groups is almost the same at the medium level. However the comparison with the day before is different. Subjects in the rural environment feel higher stress compared to previous day while subjects in urban feel almost same level of stress with previous day as shown in figure 7.

An automatic stress estimation from smart phone data is feasible as shown for example in [4]. However, in this contribution we did not further investigated an automatic stress level detection mainly due to limited sensor modalities like missing voice recordings.

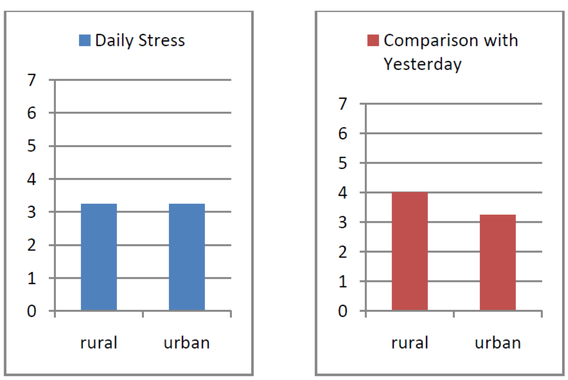

Figure 7Experienced stress level obtained from the daily survey. (Left) "How stressed have you felt today?": 1=very low, 5=very high. (Right) "How stressed have you felt today compared to yesterday?": 1=much less, 5=much more

4. CONCLUSION AND OUTLOOK

In this feasibility study, basic lifestyle parameters were measured and compared between urban and rural environments using smart phones. Five people from an urban area and five people from a rural area participated in our study during five consecutive days. An extended framework for Android smartphones was employed to obtain subjective ratings on lifestyle parameters and to collect objective sensor data. Correlation analysis between subjective data and sensor features was performed to show the potential of an automatic detection of health-relevant lifestyle parameters. It could be shown that physical activity level of subjects who live in rural environment is higher compared to the subjects in the urban area. A reasonable high linear correlation coefficient of 0.74 was found for the feature variance of x-axis acceleration. Working times were slightly increased in the rural population while commuting time was increased in the urban population. Both parameters might be estimated by GPS and/or WiFi fingerprint readings in the future. The level of social interaction in rural environment was found higher than in the urban environment. A negative correlation coefficient was found when correlating face to face communications with number of Bluetooth devices around. Participants living in the rural area showed less sleep duration than those living in the urban environment. Correlating sleep duration with a screen on/off metric resulted in a relative low correlation coefficient of 0.11. Stress level of both groups of subjects are observed similar and at mid-range.

In a follow-up study, the number of participants and the duration of the monitoring period have to be increased. The resulting extended data set will allow (1) to verify the results presented in this contribution, (2) to perform statistical test for comparing both groups, and (3) to compute parameters like participant's working time, commuting time and even stress level which were not assessed by objective markers in this contribution due to limited amount of data.

5. ACKNOWLEDGEMENT

This work was partially funded by the Co-Funded Brain Circulation Scheme Project "Pervasive Healthcare: Towards Computational Networked Life Science" (TÜBITAK Co-Circ 2236, Grant agreement number: 112C005) supported by TÜBİTAK and EC FP7 Marie Curie Action COFUND.

6. REFERENCES

[1] Nicholas D. Lane, Mashfiqui Mohammod, Mu Lin, Xiaochao Yang, Hong Lu, Shadid Ali, Afsaneh Doryab, Ethan Berke, Tanzeem Choudhury, and Andrew Campbell. Bewell: A smartphone application to monitor, model and promote wellbeing. Proceedings of 5th International ICST Conference on Pervasive Computing Technologies for Healthcare, 201, 23-26

[2] Funf Open Sensing Framework http://www.funf.org/

[3] Liu, S., Jiang, Y., & Striegel, A. Face-to-face proximity estimation using Bluetooth on smartphones. IEEE Transactions on Mobile Computing, 13:4, 2013, 811-823

[4] Amir Muaremi, Bert Arnrich, Gerhard Tröster (2013). Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep. BioNanoScience, 3:2, 2013, 172-183