Predictors of Life Satisfaction based on Daily Activities from Mobile Sensor Data

Onur Yuruten
Human Computer Interaction Group
EPFL, Switzerland
onur.yuruten@epfl.ch

Jiyong Zhang
AI Laboratory
EPFL, Switzerland
jiyong.zhang@epfl.ch

Pearl Pu
Human Computer Interaction Group
EPFL, Switzerland
pearl.pu@epfl.ch

ABSTRACT
In recent years much research work has been dedicated to detecting user activity patterns from sensor data such as location, movement and proximity. However, how daily activities are correlated to people's happiness (such as their satisfaction from work and social lives) is not well explored. In this work, we propose an approach to investigate the relationship between users' daily activity patterns and their life satisfaction level. From a well-known longitudinal dataset collected by mobile devices, we extract various activity features through location and proximity information, and compute the entropies of these data to capture the regularities of the behavioral patterns of the participants. We then perform component analysis and structural equation modeling to identify key behavior contributors to self-reported satisfaction scores. Our results show that our analytical procedure can identify meaningful assumptions of causality between activities and satisfaction. Particularly, keeping regularity in daily activities can significantly improve the life satisfaction.

Author Keywords
Analysis Methods; Ubiquitous Computing / Smart Environments; Handheld Devices and Mobile Computing

ACM Classification Keywords
H.1.2 User/Machine Systems: Human information processing

INTRODUCTION
Pervasive healthcare systems provide automated wellness monitoring [3] and activity suggestions to improve the well-being of the user. The user is equipped with various sensors, which collect information on users' metabolism, activity, location, and so on. The ever-increasing number of diseases and deaths due to inactivity [1] strongly indicate that such systems should become an indispensable component of our lives.

[1]see http://www.bbc.co.uk/news/uk-wales-politics-18876880

CHI 2014, April 26–May 1, 2014, Toronto, ON, Canada.

ACM 978-1-4503-2473-1/14/04$15.00.
http://dx.doi.org/10.1145/2556288.2557147

There are already various studies which investigate the goals that should and can be achieved through such systems, such as maintenance of physical health [1, 10, 11, 12], and providing the means for self-monitoring [5, 6, 10, 11].

Recognizing and analysing human activity patterns can be used to improve the design for healthcare systems and lifestyle recommenders. While there exists several studies that employ different techniques to recognize activity patterns and emotional states [6, 9, 13]; little has been done to understand the relations between these patterns (such as pairwise correlations between different sources of data [10]). Thus, a detailed analytical procedure is required to identify the structural relations between activities and well-being.

In this study, we are interested in two issues. First is to find the activity patterns of users using the information collected from mobile devices. Second is to investigate how daily activities are correlated to people's satisfaction from life, measured through survey data. We adopt the well-known Reality Mining dataset [2] in our study, through which we extract daily activities and apply Structural Equation Modeling (SEM) to find their relations with reported levels of satisfaction.

METHODS

Sensor Data in Reality Mining
The Reality Mining Dataset [2] is collected between 2004 and 2005 from a longitudinal study in MIT, Boston. It includes 94 users (students, professors and staff). Each participant was given a mobile phone with several pre-installed pieces of software to record various information, including call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and phone status (such as charging and idle). The Reality Mining dataset has been used in many studies. For example, Zheng et al. [13] use probabilistic reasoning techniques to discover behavior patterns and group users.

Feature Extraction
The dataset do not include annotations of user daily activity. We use *communication* information (the number of SMS and voice calls every hour), *proximity information* (the number of devices discovered in bluetooth scans every 5 minutes) and *location* information (hourly recorded as *home, work, elsewhere*) in the dataset to *estimate* the activities. For this, we consider *intrapersonal* data, i.e., we compare the proximity information of a person in a given hour with his own average

	High-Social $(P(u,t) > \bar{P}_u)$	Low-Social $(P(u,t) \leq \bar{P}_u)$
Phone On, Location = *Home*	HomeSocial	HomeRest / Sleep*
Phone On, Location = *Work*	WorkSocial	Working
Phone On, Location = *Elsewhere*	LeisureSocial	LeisureRest
Phone Off from Mid-night to 8AM	Sleep	
Phone Off from 8AM to Mid-night	PrivateActivity	

Table 1. The rules to estimate the activities for a given user u. (*): Being at home from 8PM to 8AM without any SMS or voice call action is labeled as *Sleep*)

hourly proximity information. For a given time t, $P(u,t)$ denotes the number of bluetooth devices in the proximity of the user u, and $\bar{P}_u$ denotes the average number of bluetooth devices discovered hourly by the same person. If $P(u,t) > \bar{P}_u$, we label the user in *high-social* mode, otherwise he is labeled in *low-social* mode. We summarize these activities (8 in total) in Table 1. According to our estimation method, an average student spends 7.4 hours on sleep, 6.8 hours on work, 1 hours on break in work (*WorkSocial*), 7.3 hours on leisure outside (*LeisureRest* and *LeisureSocial*), and 1.5 hours on other activities (HomeSocial, HomeRest, PrivateActivity). These numbers are similar to the findings in 2012 Sodexo University Lifestyle Survey report [2].

We also include four *communication-related* features: the number of SMS, number of phone calls, *proximity* (proximity count and proximity time) information. We also notice that users' activities are very different between weekdays and weekends, so we calculate them separately. Thus we obtain 24 features for activities on weekdays and weekends for each user as shown in Table 2.

Id	Activity Features
1-2	HomeSocial {weekday, weekend}
3-4	HomeRest {weekday, weekend}
5-6	Sleep {weekday, weekend}
7-8	WorkSocial {weekday, weekend}
9-10	Working {weekday, weekend}
11-12	LeisureSocial {weekday, weekend}
13-14	LeisureRest {weekday, weekend}
15-16	PrivateActivity {weekday, weekend}
17-18	BluetoothDeviceCount {weekday, weekend}
19-20	BluetoothDeviceTime {weekday, weekend}
21-22	VoiceCallCount {weekday, weekend}
23-24	SMSCount {weekday, weekend}

Table 2. Activity Features.

Lastly, we represent the regularity of user activities through their *entropies*. The entropy of a feature x can be calculated as:

$$\mathcal{H}(x) = -\sum_{t\in[1,24]}\sum_{c\in\mathcal{C}_x} p(c|t)log(p(c|t)), \qquad (1)$$

where t denotes the time in hours. $\mathcal{C}_x$ is a generic notation for the set of available items of the given feature x. For instance, for activity entropy $\mathcal{H}(x = activity)$, $\mathcal{C}_x$ is the set of all possible activities. Then, $p(c|t)$ would denote the probability of having activity c at time t. A person with high activity entropy would have irregular amounts and distribution of activities while he/she participated to the longitudinal study. We compute entropies for *activity*, *social time*, *location* and *proximity* for each user for weekdays and weekends respectively, providing us with *8* regularity features as shown in Table 3.

Id	Regularity Features
25-26	ActivityEntropy{weekday, weekend}
27-28	LocationEntropy{weekday, weekend}
29-30	ProximityEntropy{weekday, weekend}
31-32	SocialEntropy{weekday, weekend}

Table 3. Regularity Features.

Survey Data

There are *25* survey questions in the dataset, *10* of which are *self-reported* measures of happiness, health and travel frequency of the users (See Table 4). In this work we use such self-reported satisfaction information to represent users' happiness. These fields have different ranges: question with $Id = 42$ ranges between 1-5; question with $Id = 41$ ranges between 1-4; and the rest range between 1-7. To obtain a unified interpretation, we scaled their values to a common range, i.e., 1-5. We have performed a linear scaling, which retains the information of the original values.

Id	Survey Questions
33	I am satisfied with my experience at MIT thus far
34	I am satisfied with my current social circle
35	I feel I have learned a lot this semester
36	I am satisfied with the content and direction of my classes and research this semester
37	I am satisfied with the support I received from my circle of friends
38	I am satisfied with the level of support I have received from the other members in my group
39	I am satisfied with the quality of our group meetings
40	I am satisfied with how my research group interacts on a personal level
41	Have you been sick recently?
42	Have you travelled recently?

Table 4. Satisfaction/wellness questions

Identifying Factors

We apply Principal Component Analysis (PCA [8]) on the features to group them into *factors* (expressed in terms of eigenvalues and eigenvectors). We choose the eigenvalue threshold as *1* to determine the number of factors, and the factor loading threshold as *0.55* in order to include the features for further analysis. We use the IBM SPSS software

[2]http://uk.sodexo.com/uken/media-centre/press-releases/university-lifestyle.asp

		Component					
		1	2	3	4	5	6
Proximity Entropy	**weekdayProximityEntropy**	**.920**	-.034	-.031	-.094	.145	-.085
	weekendsBluetoothTime	**.858**	-.270	.110	-.010	-.091	.013
	weekendProximityEntropy	**.858**	.048	-.077	-.089	-.025	-.201
	weekdaysBluetoothTime	**.805**	-.166	.001	.060	.093	.100
	weekendsBluetoothCount	**-.750**	-.071	.020	-.056	.072	.117
	weekdaysBluetoothCount	**-.750**	-.188	-.001	-.130	-.080	.140
Communication Activities	**weekendsVoiceCount**	-.006	**-.853**	.137	.052	.010	.127
	weekdaysSMSCount	-.015	**.841**	-.093	.172	.064	-.048
	weekdaysVoiceCount	.050	**-.819**	.156	.110	.114	.018
	weekendsSMS	.013	**.798**	.006	.148	.101	-.121
	weekendActivityEntropy	.325	**.567**	-.358	-.302	.276	.120
	weekendSocialEntropy	*-.372*	*-.455*	*.411*	*.217*	*-.017*	*-.265*
	weekdayHomeRest	*-.148*	*.291*	*.167*	*.021*	*-.145*	*.014*
Leisure and Sleep Activities	weekendHomeSleep	-.078	.123	**-.798**	-.044	.089	-.192
	weekdayLeisuring	.064	-.029	**.763**	-.376	.266	.091
	weekendLeisuring	.050	-.018	**.758**	-.394	.231	.207
	weekdayHomeSleep	-.040	.116	**.687**	.449	-.287	.187
	weekdayActivityEntropy	-.273	-.334	**.579**	.322	-.186	-.079
	weekendLeisureSocial	*-.065*	*.312*	*-.440*	*-.021*	*-.097*	*.271*
	weekdayLeisureSocial	*.327*	*.113*	*-.421*	*-.288*	*.082*	*.227*
Working Activities	**weekdayWorking**	.120	.072	-.113	**.865**	.259	.124
	weekdayWorkSocial	-.129	.108	.106	**.789**	-.015	.042
	weekendWorking	.272	-.017	.076	**.742**	.218	.078
	weekendWorkSocial	.099	.051	.032	**-.553**	-.059	.069
	weekendHomeRest	*.070*	*-.140*	*-.125*	*.178*	*-.162*	*-.134*
Social Entropy	weekendPrivateActivity	.016	.034	.016	-.039	**-.799**	.030
	weekdayPrivateActivity	-.126	.005	.012	-.334	**-.707**	-.036
	weekdaySocialEntropy	.461	.182	-.195	-.163	**.565**	.060
	weekendHomeSocial	-.078	.020	.159	.199	**.551**	-.373
Location Entropy	**weekdayLocationEntropy**	-.223	-.226	.186	.058	-.031	**.839**
	weekendLocationEntropy	-.309	-.189	.166	.221	-.011	**.808**
	weekdayHomeSocial	*.113*	*.239*	*-.065*	*-.156*	*-.400*	*.441*

Figure 1. Principal Component Analysis for Activity Features.

		Component	
		1	2
Social Life Satisfaction	SAT.GroupInteractPersonal	**.848**	-.223
	SAT.SupportFromGroupMembers	**.786**	-.151
	SAT.GroupMeetings	**-.738**	.035
	SAT.SupportFromFriends	**.735**	-.056
	SAT.SocialCircle	*-.457*	*.033*
Research Satisfaction	SAT.Learning	-.399	**.707**
	SAT.ResearchContentAndDirection	-.277	**.682**
	SAT.Overall.MIT	-.270	**.634**
	Health	*-.043*	*-.513*
	TravelFrequency	*-.111*	*-.473*

Figure 2. Principal Component Analysis for Survey Data.

and apply the varimax rotation (an orthogonal rotation) in the factor analysis. We identify 6 *factors* from the activity features, and 2 *factors* from the survey data. We name the factors with respect to the feature with the highest positive loading, as conveyed in Figures 1 and 2. More precisely, two factors are satisfaction-related (*Social Life Satisfaction* and *Research/Study Satisfaction*), and another three are regularity-related (*Social Entropy, Location Entropy* and *Proximity Entropy*) and the remaining three are related with the activity patterns (*Leisure and Sleep*, *Working Activities*, and *Communication Activities*). We discard some features (*WeekdayHomeRest*, *WeekendHomeRest*, *WeekendSocialEntropy*, *WeekdayLeisureSocial*, *WeekendLeisureSocial*, *WeekdayHomeSocial*, *Health*) since their loadings are below *0.55*.

Structural Equation Modeling

To understand how these activity features affect self-reported satisfaction, we use Structural Equation Modeling [7], which can be used both to explore and confirm hypotheses of causal assumptions between groups of features, and model noises in the data with latent (unobserved) variables. To our knowledge, there is only one study that uses SEM for daily activity analysis - specifically, for predicting sequence of activities based on commute data [4]. The dataset of that study includes solely self-reported activities and their durations. In contrast, the Reality Mining dataset was collected using modern sensor technology.

We have followed commonly accepted thresholds for factor analysis[3]: as shown in Figures 1 and 2, each factor has at least 2 features with factor loading larger than 0.7.

RESULTS AND ANALYSIS

The structural model fit for our hypothesis is shown in Figure 3. The model goodness-of-fit indices ($\chi^2 = 1160.5$, $df = 473$, $p < 0.05$, $RMR = 0.008$), and R^2 values ($R^2 > 0.1$ for all) surpass their recommended values. In this model, we have drawn paths from the activity related factors (Entropies, leisure, sleep, work, and communication activities) to the satisfaction-related factors. The analysis conveys three interesting causal assumptions:

- *Social Entropy - Leisure and Sleep - Social Life Satisfaction:* The increase in social activity regularities (i.e., the decrease of *Social Entropy*) improves both *Leisure and Sleep*, and *Social Life Satisfaction*. Furthermore, *Leisure and Sleep* also has direct positive influence on *Social Life Satisfaction*. Thus, we say that *Social Entropy* has an amplifying effect. To illustrate this, we select top ten regular users and top ten irregular users with respect to the feature of *social entropy*, and compare their satisfaction levels. We observe that in average the regular users report 40.74% higher satisfaction score (significant, $p = 0.023$) in the survey question with $Id = 34$ than the irregular users.
- *Working Activities - Leisure and Sleep - Social Life Satisfaction: Working Activities* have an amplifying effect on *Leisure and Sleep* and *Social Life Satisfaction*, but with a different interpretation: While both *Working activities* and *Leisure and Sleep* positively influence *Social Life Satisfaction*, working activities have negative influence on leisure and sleep. This implies that spending more time at work lowers the time for sleep and other activities. However, this analysis does not exactly show how to compute an equilibrium between work and leisure and sleep activities.

[3]see http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/thresholds for a summary of thresholds

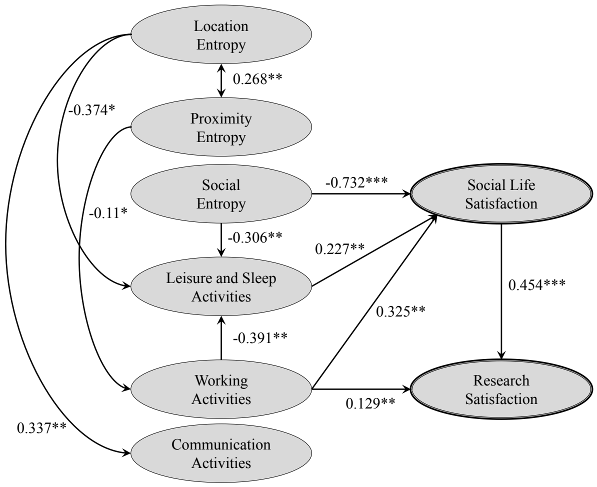

Figure 3. The model fitted with SEM. The values on the directed paths denote the standardized regression weights of the model. For example, when *Leisure and Sleep Activities* goes up by one standard deviation, *Social Life Satisfaction* also goes up by 0.227 standard deviations. The paths with significance levels $p < 0.1$, $p < 0.05$, and $p < 0.001$ are marked with *, ** and *, respectively. For brevity, we omitted the features of the factors and the paths that do not have statistical significance.**

- *Working Activities - Social Life Satisfaction - Research Satisfaction: Research Satisfaction* is positively influenced by both *Working Activities* and *Social Life Satisfaction*. Similar with the previous observation, the *Working Activities* factor has an amplifying effect.

Other regularity-related factors (location, bluetooth proximity) are crucial to *Leisure and Sleep*, *Communication*, and *Workplace Activities*. Thus they indirectly regulate satisfaction: the lower the entropy is, the higher satisfaction with research and social life a user would have.

CONCLUSION AND FUTURE WORK

We have performed a statistical analysis on the Reality Mining dataset to identify the predictors of life satisfaction. Our results show that our method is useful for estimating user activities, and for identifying meaningful relations between activities and satisfaction. More specifically, our analysis shows that work, leisure and sleep activities, and regularities in the daily activities have both direct and indirect influences over the reported levels of satisfaction.

These findings can guide us toward better designs for lifestyle recommender systems. Our long-term research goal is to design a lifestyle recommender system that provides accurate personalized suggestions on daily activities to improve the wellness of its users. In future we will perform advanced time-series analysis on the relationship between daily activities and life satisfaction. Furthermore, the Reality Mining dataset provides a limited amount of information for our purpose. We will launch our own user studies to obtain a more dedicated dataset with continuous sensor measurements on user mood and physical activities.

REFERENCES

1. Denning, T., Andrew, A., Chaudhri, R., Hartung, C., Lester, J., Borriello, G., and Duncan, G. Balance: towards a usable pervasive wellness application with accurate activity inference. In *Proceedings of the 10th workshop on Mobile Computing Systems and Applications*, no. 5 (2009).
2. Eagle, N., Pentland, A., and Lazer, D. Inferring social network structure using mobile phone data. In *PNAS*, vol. 106 (2009), 15724–15278.
3. Korhonen, I., Pärkkä, J., and Gils, M. V. Health monitoring in the home of the future. *IEEE Engineering in Medicine and Biology Magazine 22*, 3 (2003), 66–73.
4. Kuppam, A., and Pendyala, R. A structural equations analysis of commuters' activity and travel patterns. *Journal of Transportation 28*, 1 (2012), 33–54.
5. Li, I., Dey, A. K., and Forlizzi, J. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proc. 13th international conference on Ubiquitous computing* (2011), 405–414.
6. McDuff, D., Karlson, A., Kapoor, A., Roseway, A., and Czerwinski, M. Affectaura: An intelligent system for emotional memory. In *SIGCHI Conference on Human Factors in Computing Systems* (2012), 849–858.
7. Pearl, J. *Causality: Models, Reasoning, and Inference*, vol. 29. Cambridge: MIT Press. ISBN 0-521-77362-8, 2000.
8. Pearson, K. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2*, 11 (1901), 559–572.
9. Suryadevara, N. K., Quazi, T., and Mukhopadhyay, S. C. Smart sensing system for human emotion and behaviour recognition. *Perception and Machine Intelligence 7143* (2012), 11–22.
10. Tollmar, K., Bentley, F., and Viedma, C. Mobile health mashups: Making sense of multiple streams of wellbeing and contextual data for presentation on a mobile device. In *6th International Conference on Pervasive Computing Technologies for Healthcare* (2012), 65–72.
11. Tsai, C. C., Lee, G., Raab, F., Norman, G. J., Sohn, T., Griswold, W. G., and Patrick, K. Usability and feasibility of pmeb: a mobile phone application for monitoring real time caloric balance. *Mobile Networks and Applications archive 12* (2007), 173–184.
12. Zhang, K., Pi-Sunyer, F., and Boozer, C. Improving energy expenditure estimation for physical activity. *Medicine and Science in Sports and Exercise 36*, 5 (2004), 883–889.
13. Zheng, J., and Ni, L. M. An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ACM New York (2012), 153–162.