An Exploration of Daily Routine Modeling based on Bluetooth and GSM-data

Ulrich Steinhoff, Bernt Schiele
TU Darmstadt, Germany

Abstract

We explore GSM & Bluetooth (BT) data to capture daily routines and social environments. With focus on the individual, non-networked user, showing the complementarity of these modalities. We present a realistic real-life dataset over 22 weeks and analyze probabilistic topic models to automatically discover typical situations of everyday life.

1. Introduction

Daily routine modeling is interesting for applications, e.g. in industry or health care. We record a realistic 22-weeks dataset, analyze if BT-&GSM-data contains interesting information for daily routine modeling and apply probabilistic topic models to automatically learn this structure.

Related Work. Many works exist on user context by location, e.g. [1] uses GPS to infer location, destination and transportation mode, [2] uses BT/GPS information to extract social structure and significant places. [3] recorded BT(& GSM) of 100 users to infer home/work/other status from GSM and analyze social networks. [4] uses topic models on this dataset to find typical location-driven user routines and interaction patterns on a daily scale. In contrast to previous work which mainly analyzes the relations between known participants we focus on BT/GSM environments of an individual without any annotation of other devices prior to automatic daily routine discovery. Interestingly, topic models also allow to extract daily routines using a different sensor modality, namely on-body accelerometers [5].

2. Dataset

To gain insight into a typical current BT/GSM environment, we recorded 22 weeks of unaltered single-user data, including working routine, two business trips, several project meetings, leisure activities, and three holiday travels abroad. We used a low-cost mobile phone, supporting J2ME API's for BT and Cell Broadcast Service (CBS, for current cell ID) and allowing long-term logging with a scanning interval of 30s at runtimes over 24h. We consider only parts with valid CBS data, resulting in 1933h of data (over 220k Scans).

GSM cells. We encountered 505 GSM cells, which showed strong exponential distribution. The 4 top cells account for 36.8%, 15.3%, 14.6% and 9.3% of GSM records; only 23% resp. 9% of cells were seen for $\geq$5 resp. $\geq$30 minutes and 14% of cells in at least 3 different days. To cope with the large amount of non-significant cells and subsume locations out of everyday life (e.g. seminar, project meeting), we subsume all cells seen less than 5h or in less then 7 days as 'other', accounting for 16% of the time overall.

Bluetooth (BT). BT devices are a potential rich but noisy source for fine-grained location information as well as for the social environment given by personal devices. In average 1.6 BT devices are in one scan and 86.5% of 10min windows contain a device. About twice as many devices are seen during working hours as at home.

We identified manually five major classes by appearance patterns in our dataset: Work environment contains two classes; *Work Static*, with the user's desktop computer as most prominent example and *work social* with the part of personal devices of coworkers (laptops and phones), that is also seen at other places, e.g. canteen or external project meetings, not only at work place. *Work Social* indicates a working situation independently of the actual location. In home environment the flatmates' and neighbors' devices represent a *Home class* of 'human landmarks' similar to static devices. Even though the devices are mobile phones and laptops, they are only seen in the flat. We found also frequently spotted devices, especially occurring at home *and* at work, which fall into a *Personal Device class*. As the phone itself is not included in the scans, the subject's laptop is the main device of this category. 6126 unique devices were observed in total, of which 69% are seen in $\leq$ 2 scans, 88% less then 10 scans and 93% in less then 3 different days. To cope with the large number of random encounters and non-significant devices, we decide to subsume all BT devices seen in less than 7 different days or for less than 5h in an *'Other' class*, which shows a good tradeoff between importance and number of devices (24 devices left in above classes, accounting for 87,1% of BT records). Other devices of interest, e.g. of the subject's friends are underrepresented and only occasionally set visible, and thus in the'other' class. However, the presence of many 'other' devices hints that the user is situated in a public space, transit or generally non-everyday situation.

Overall we conclude that the dense BT data contains a social dimension as well as inherent location context due to reappearing, user-dependent sets of devices. Identifying those sets by hand is a tedious task and undesirable for a normal user, thus we explore in the following section a way to obtain meaningful sets automatically.

3. Automatic Discovery of Daily Routines

Probabilistic Topic Models (PTM). To apply topic models, we transform the recorded data to a 'bag-of-words' rep-

Table 1. Specific devices/cells for topics shown in Fig.1

Topic	most important devices/cells	% of activations
BT1	user's desktop, officemate's phone&laptop	25.6
BT2	phone&laptop of flatmate1	19.0
BT3	devices of other colleagues	14.5
BT4	personal laptop & flatmate2's laptop	30.2
BT5	'other' BT devices	10.7
GSM1	Home (top1), Home (top3)	38.7
GSM2	'other' cell	16.0
GSM3	Home (top2)	14.2
GSM4	Innercity (ic) cells + home/ic + uni/ic	5.6
GSM5	University cells (top1-3)	25.8

Table 2. Situations labeled in in Fig.1

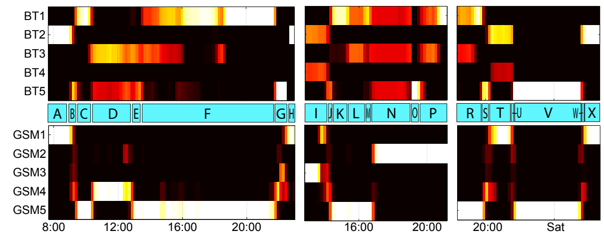
	situation	dominant *BT*	topics *GSM*		situation	dominant *BT*	topics *GSM*
A	home	2	1	*M*	office	1	5
B	commuting	5	4(+all)	*N*	2h car drive	1,3,5	2
C	office	1	5	*O*	exploring site	5	2
D	external meeting	3,5	4	*P*	dinner w/ coll.	1,3	2
E	canteen	3,5	5	*R*	office	1,3	5
F	office	1,3	5	*S*	commuting	5	4
G	walk home	5	all	*T*	home	2,4	1
H	home	2	1	*U*	bicycling	(5)	1,2,4,5
I	home	2,4	3,1	*V*	club near uni	5	5
J	commuting	5	4(+all)	*W*	bicycling	(5)	1,2,4,5
K	office	1	5	*X*	home	2	1
L	in nearby cafe w/ colleagues	1,3	5				

resentation by building histograms on detected BT devices, presence of 'other' devices and/or current cell in 20 minutes time slices (40 scans), which shows a good tradeoff between amount of scanned devices per timeslice and granularity. We use these histograms as documents for a Latent Dirichlet Allocation (LDA) topic model in an implementation of [6], varying hyperparameters and number of topics. In contrast to unsupervised clustering methods, PTM allow a representation of documents/timeslices as a co-activation of underlying topics instead of assignment to a single cluster.

Topic activations for BT and GSM. Extracting 5-6 topics using hyperparameter α values near or greater 1, we can typically obtain easily interpretable topics for daily life's BT environment, resembling the manually determined classes. While with lower α devices are more shared between topics less co-activated within timeslices, higher α favors topics with just single devices picked and more co-activation of topics. Thus lower α lead to results that are more similar to unsupervised clustering on the histograms. Using higher number of topics the most frequent devices separate to own topics, alone or for frequent combinations. On GSM we typically obtain topics correlating to Home/Uni/Innercity/other, with often multiple topics for home due to seldom switching of cells there.

Table 1 shows the characteristic content for 2 representative sets of 5 topics built on BT resp. GSM data. Fig.1 presents the relative strength of topic activations for 3 examples, with situations described in Table 2. Fig.1(a) shows a normal working day (with a project meeting in inner city), Fig.1(b) travelling to a seminar with colleagues, and Fig.1(c) going out near university.

Noteworthy are the not clear-cut topic BT4, joining a flatmate's device with the personal laptop due to long concurrences, and GSM4, which groups innercity cells with 2 ambiguous cells. GSM indentifies a coarse location context of the user, which is for the most part a good indicator for the user's work/leisure situation, but only additional BT topic activation can tell situation V apart from the other stays in university area, as we see activation of the 'other' topic BT5 (typical for crowded environments and during commuting) and no co-activation of work related topics. The other way around, BT5 is often exclusively activated for a stay in a crowded place as well as for commuting, but co-activation of several GSM topics is only seen for the latter. In general BT topics offer a possibility to determine work-related situations, as office F and seminar P, where for the latter GSM shows that it is a non-everyday situation.



(a) working day w/ meeting (b) traveling day (c) work/fri night

Figure 1. Topic activations on BT/GSM for 3 examples.

4. Discussion & Conclusion

While BT devices are visible for large parts during everyday live nowadays, we also observe that BT-environments and therefore the discovered structure is inherently unstable and biased towards a small fraction of the observed devices.

Long time periods spend at home and office are posing difficulties to a statistical approach (e.g. the devices of one officemate are typically grouped with static office devices). As topic models are based on co-occurrence within a timeslice, they are not well suited for GSM traces with only the current cell. In our dataset many handovers occur at university, but home shows countinous cell attachment. Thus few timeslices obtain co-occurrences between different home-related cells.

We conclude that interesting structure of an individual's daily life is reflected in BT and GSM environments. It is possible to discovery this structure in an unsupervised fashion, at least partially, by incorporating more devices than a user might identify manually. Even in unknown places or for cells in which several activities take place, BT-environments have the potential to identify the user's situation. The extracted topics could be also useful as an intermediate representation for supervised high level activity recognition in conjunction with additional sensors (e.g. fine-grained location, inertial).

References

[1] L. Liao, D. Fox, and H. Kautz, "Extracting Places and Activities from GPS Traces Using Hierarchical CRFs," *IJRR*, vol. 26, 2007.

[2] B. Adams, D. Phung, and S. Venkatesh, "Sensing and using social context," *ACM TOMCCAP*, vol. 5, no. 2, 2008.

[3] N. Eagle and A. Pentland, "Reality mining: sensing complex social systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, 2006.

[4] K. Farrahi and D. Gatica-Perez, "Discovering human routines from cell phone data with topic models," in *ISWC*, 2008.

[5] T. Huỳnh, M. Fritz, and B. Schiele, "Discovery of Activity Patterns using Topic Models," in *UbiComp*, 2008.

[6] M. Steyvers and T. Griffiths, "Probabilistic topic models," in *Handbook of Latent Semantic Analysis*, 2007.