URB 03 Inferring Everyday Mobility States using GSM & Wi-Fi Traces from Mobile Phones

URB 03.1 People

- Principal Investigator: Deborah Estrin, Mark Hansen, Jeff Burke
- Faculty: Deborah Estrin, CS, UCLA; Mark Hansen, STAT, UCLA
- Staff: Jeff Burke, REMAP, UCLA
- Graduate Students: Min Y. Mun, CS, UCLA

URB 03. 2 Overview

Inferring mobility states such as being stationary, walking, or driving is critical for many applications in transportation studies, urban planning, health monitoring and epidemiology.

Our focus is on building a pervasive mobility classification system using mobile phones with the goal of large deployments. More specifically, the classification models should provide the following:

• *Low Processing Complexity*: Although mobile phones have advanced drastically in functionality and capability in the past few years, they still have limited resources in terms of CPU and memory (RAM). The classification models for mobile-based applications should be lighter weight and supported alongside everyday applications in a robust and resilient manner.

• *High Energy Efficiency*: People carry mobile phones throughout the day. Running power-expensive applications will significantly reduce the lifetime of mobile phones and burden users; users need to recharge their battery in the middle of the day; otherwise, they would not be able to use their phones for communication usages as well as application services.

• *High User-time Coverage*: To recognize users activities in everyday situations, systems must have ubiquitous usertime coverage. GPS positioning is available as little as 5% of a typical person's day, while GSM and Wi-Fi coverage is available throughout the day [1].

URB 03.3 Approach

GPS positioning [2-6], external geoindexes such as map information, commercial and custom devices such as pedometers [7,8] and body sensors [9-12] have been widely used to determine people's fine-grained activities. Unfortunately, it is not always practical to deploy mobility-based sensing applications [13] using such activity classifiers on a large scale due to their cost, complexity, obtrusiveness, coverage and energy profile. In particular, GPS-based mobility characterization raises many issues such as spotty coverage and battery drainage that makes it inadequate to meet all application goals.

Mobile phones are widely deployed and people carry them everyday and all the time, which indicates that the mobile phone is one of the most suitable devices to use in recognizing people's everyday physical activities.

We propose a new mobility classification method using GSM and Wi-Fi traces that are available on many commercial mobile phones. We adopted C.4.5 Decision Tree as our inference model. It is light and simple, which makes our model suitable to be used in mobile phones for its low processing complexity. Sampling GSM and Wi-Fi data uses relatively low energies, supporting high-energy efficiency. In addition, GSM and Wi-Fi coverage is nearly ubiquitous, which provides a solution to high user-time coverage. Exploring how mobility classification can be performed using GSM and Wi-Fi observations contributes to profiling unconstrained mobility states throughout a typical day on a large scale for the participatory sensing applications including our ongoing application. Our Personal Environmental Impact Report (PEIR) project uses automated mobility-annotated location traces as an input to models of hazard exposure and environmental impacts as described previously. The work described here could enable PEIR to offer its participants personalized information on environmental interaction throughout the day in an unobtrusive way by sharing only coarse-grained location.

URB 03.4 System(s) Description and/or Experiments

GSM and Wi-Fi

GSM is the most popular standard for mobile phones in the world. Information from the cell ID provides a rough indication of a person's position. Features derived from this information, such as the number of changes in the associated cell IDs for certain duration, can substitute for the speed values from GPS data that has been widely used to identify mobility states or activities. Unfortunately, given the nature of large GSM cell sizes, a person's locations and mobility states within a single cell cannot be identified well from GSM data alone. Therefore we explore the use of data from networks with smaller cell sizes, such as Wi-Fi. We found that the accuracy of our model increases by 10% when we employed both GSM and Wi-Fi information compared to using GSM alone.

In addition, GSM and Wi-Fi coverage is nearly ubiquitous, providing high user-time coverage, while GPS positioning do not work in many situations. And sampling GSM and Wi-Fi uses relatively low energy and provides high-energy efficiency as shown in Table 1.

Activity	Power (Watts)
Phone Idle	0.054
GSM Sampling	0.056
GPS, Wi-Fi Sampling	0.23
GPS Indoor Sampling	0.407
GPS Outdoor Sampling	0.38
Accelerometer Sampling	0.11

Feature Selection

A large portion of mobiles including our experimental device, Nokia N 95, does not have access to cell tower information for cells other than the one it is associated with. In our studies, we use single connected cell ID information to extract features discriminating the mobility states. The intuition is that users see more cells for certain duration, as they move faster. The following features are extracted from GSM traces, where w is the sliding window size:

Number of Unique Cell IDs (C unique,w) is the number of unique cell IDs to which the phone was connected for w second segment.

- *Residence Time in a Cell Footprint* (C residence) is the duration a user spent in the cell that the phone was associated with. Note that C residence is different from the other two features because we divide data into segments by cell IDs. Thus w value can vary.
- Although Wi-Fi data has many features satisfying our design requirements, it should be carefully used due to it's limitations. In the same area, users may see various Wi-Fi APs at different times. However, a Wi-Fi AP must be visible for the greatest amount of time among other APs in view while being stationary at one place and we call this AP '*dominant Wi-Fi AP*.' "Switch dominance" occurs when users move from one place to another. As users move faster, this would happen more frequently and the duration of the most dominant Wi-Fi AP would get lower. The followings are features extracted from Wi-Fi traces.

Duration of Dominant Wi-Fi Access point in View (WF dominant,w) is the amount of time that the most dominant Wi-Fi AP in each segment is seen.

- *Proportion* of *Duration of Dominant Wi-Fi Access point in View* (WF dominant_proportion,w) is equal to (WF dominant,w /V) x 100 where V is the number of valid points during w second segment.

Building and Evaluating a Model
We computed the four selected features for each data point. We adopted the C.4.5 Decision Tree as our inference model for it's simplicity, which enables our model to provide low processing complexity. It also outperforms other inference models such as Bayesian Network, Support Vector Machine (SVM) and Conditional Random Field (CRF) when features are computed over segments. Ten-fold cross validation method is used for evaluation.

Due to different GSM and Wi-Fi densities in various areas, it is important to study whether we need an environment-specific model, and whether our classifier can be effectively applied to new users without additional training procedures, in addition to further evaluating the general performance of our model. We have two data collection scenarios. First, the author collected data in five differently characterized neighborhoods, Wilshire in Westood (Wilshire), Palms, UCLA, Marina Del Rey (MDR) and East Culver City (E.Culver), chosen based on the Southern California Association of Governments (SCAG) data; for each area, the data collector conducted the three mobility states for twenty minutes each. Second, sixteen individuals, eight males and eight females between the ages of 20-45, gathered one hour data, twenty minutes for each of the three mobility states.

URB 03.4 Accomplishments

General Results
Table 2 summarizes our results. We use the commonly used matrices to evaluate our model: precision (true positive/(true positive + false positive)) and recall (true positive/(true positive + false negative)) confusion matrices. Precision is the percentage of correct predictions and recall is the percentage of correctly identified ground truth cases. Accuracy (true positive/total number of data points) is the percentage of correctly classified data points. To better understand the performance results of our model, we also built mobility classification models using GSM, Wi-Fi and GPS alone.

First, our model successfully identifies the coarse-grained mobility states with overall accuracy of 80% (precision: 80%, recall: 80%). This accuracy is lower than the one of the GPS-based model, 92% (recall: 92%, precision: 91%). But, as mentioned earlier, sampling GPS is accompanied by many drawbacks.

Second, Wi-Fi beacons with smaller cell sizes and GSM data with larger cell sizes are complementary. The classifiers using GSM and Wi-Fi data alone identify mobility states 70% and 68% correctly. The accuracy increases by 10-12% when the combination is used.

	Being Stationary		Walking		Driving		All	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
GSM,Wi-Fi	66%	92%	84%	66%	90%	83%	80%	80%
GSM	76%	71%	59%	71%	80%	69%	72%	70%
Wi-Fi	76%	61%	65%	61%	70%	84%	70%	69%
GPS	82%	93%	94%	91%	99%	91%	91%	92%

Table 2. Precision and Recall Confusion Matrices of The Classifiers using Various Sensors

Impact of Environment
Our model can provide better performance by employing the model using only GSM data when Wi-Fi beacons have no coverage or are too sparse. Generally, adding Wi-Fi features to the model improves the accuracy as explained in the previous section. For example, as shown in Figure 1, the LOAO (Leave One Area Out) based model using only GSM data works poorly in characterizing stationary states in UCLA area. Our model built by using the LOAO method increases the values to nearly 80% by taking advantage of Wi-Fi. But, if the technique is used when Wi-Fi densities are too low, computed feature values can be numerically distant from the rest of the data and it could rather degrade the model. As seen in Figure 2, the performance of our model improves and achieves 82% overall accuracy by adopting the GSM-based method when necessary.

User Variation

To evaluate the scalability of our model, we built the model with the data set gathered from one user and tested it with each of sixteen new users whose data is not in the training data set. We achieve a trimmed average accuracy of 78% and the results are promising. Certain might be unique and a training set that has a broad range of activity styles and environmental settings are necessary for a generalized classifier to work properly.

URB 03.5 Future Directions

The current work explores only the first step required to build a pervasive mobility classification system using mobile phones for the large deployment. There are numerous potential methods left unexplored that may be better suited for our system. One example is using accelerometer data. Accelerometer data has been used to recognize activities having similar speed and acceleration features. In addition, sampling accelerometer data is energy-inexpensive. Thus, accelerometer data has much potential to improve the performance of our model by disambiguating being stationary and walking states, where most of the errors occurred from our model because the two states often have similar GSM and Wi-Fi feature values, while satisfying our design requirements. Our future work involves exploring how our mobility classification system would improve with using accelerometer data.

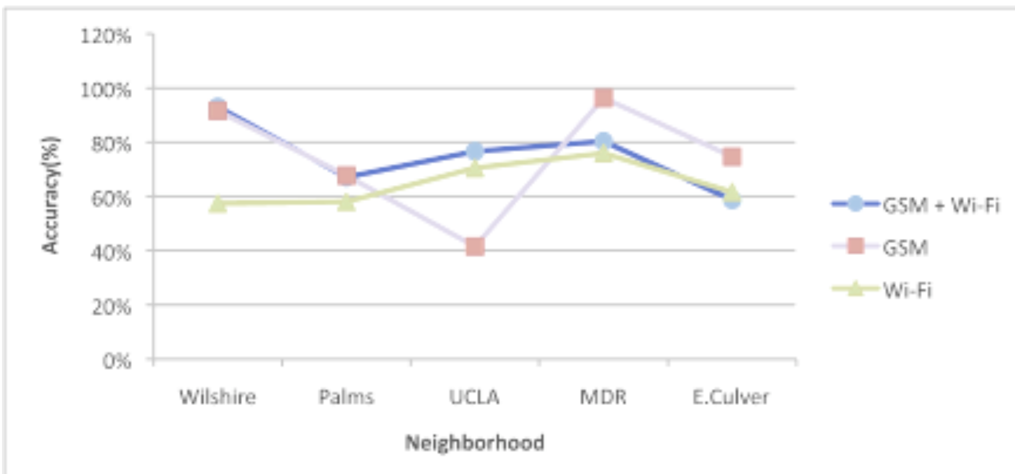

Figure 1. Leave-One-Area-Out Testing

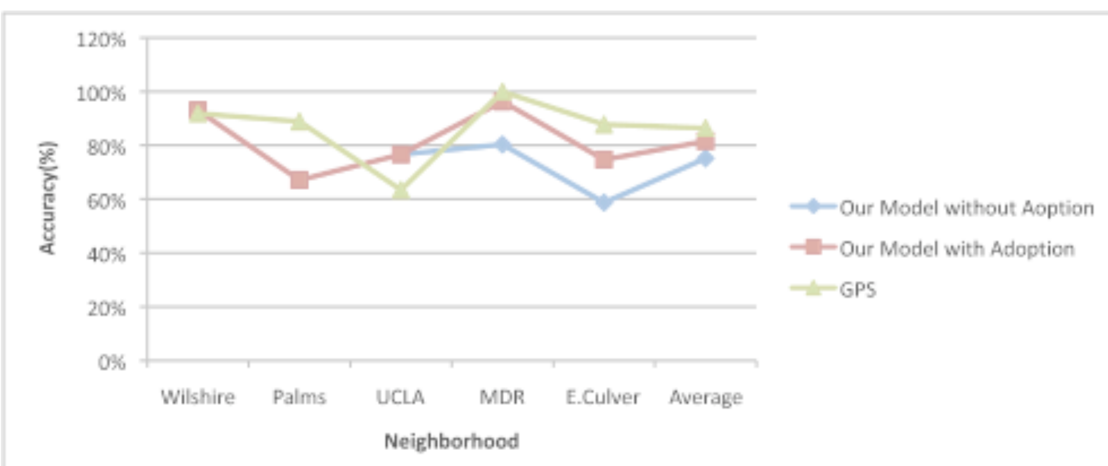

Figure 2. Comparison between Our Model with and without Adopting the GSM-based Model When Wi-Fi AP Densities are Too Sparse, and the GPS-based Model

URB 03.6 External Research Partnerships

Nokia Research Group(current)

URB 03.7 References

[1] A.LaMarca, Y.Chawathe, S.Consolvo, J.Hightower, J.Scott I.Smith, T.Shon, J.Howard, J.Hughes, F.Potter, J.Tabert, P.Powledge, G.Borriello, and B.Schilit. Place lab: Device positioning using radio beacons in the wild. In *Pervasive 2005*, Munich, Germany, 2005.

[2] L. Liao, D. Fox, and H. Kautz. Extracting places and activities from gps traces using hierarchical conditional random fields. *Int. J. Rob. Res.*, 26(1):119–134, 2007.

[3] L. Liao, D.J. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. *Artif. Intell.*, 171(5-6): 311–331, 2004.

[4] L.Liao, D.Fox, and H.Kautz. Location-based activity recognition using relational markov networks. *IJCAI*, 2005.

[5] D. Patterson, L. Liao, D. Fox, and H. Kautz. Inferring high-level behavior from low-level sensors. pages 73–89. 2003.

[6] Y.Zheng, L.Liu, L.Wang, and X.Xie. Learning transportation mode from raw gps data for geographic applications on the web. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 247–256, New York, NY, USA, 2008. ACM.

[7] Omron. http://www.omronhealthcare.com.

[8] Inc. Apple Computer. Nike+ ipod sport kit http://apple.com.

[9] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive*, pages 1–17, 2004.

[10] S. Crouter, P. Schneider, M. Karabulut, , and D. Bassett. Validity of 10 electronic pedometers for measuring steps distance and energy cost. In *Medicine & Science in Sports & Exercise*, 2003.

[11] N. Kern, B. Schiele, and A.Schmidt. Multi-sensor activity context detection for wearable computing. In *In Proc. EUSAI, LNCS*, pages 220–232, 2003.

[12] C. Randell and H. Muller. Context awareness by analyzing accelerometer data. In *ISWC '00*, page 175, Washington, DC, USA, 2000. IEEE Computer Society.

[13] J.Burke, D.Estrin, M.Hansen, A.Parker, N.Ramanathan, S.Reddy, and M.B.Srivastava. Participatory sensing. 2006.