Fusing On-Body Sensing with Local and Temporal Cues for Daily Activity Recognition

Zack Zhu
Wearable Computing Lab
ETH Zurich, Switzerland
zack.zhu@ife.ee.ethz.ch

Ulf Blanke
Wearable Computing Lab
ETH Zurich, Switzerland
ulf.blanke@ife.ee.ethz.ch

Alberto Calatroni
Wearable Computing Lab
ETH Zurich, Switzerland
alberto.calatroni@ife.ee.ethz.ch

Oliver Brdiczka
Palo Alto Research Center
Palo Alto, USA
brdiczka@acm.org

Gerhard Tröster
Wearable Computing Lab
ETH Zurich, Switzerland
troester@ife.ee.ethz.ch

ABSTRACT

Automatically recognizing people's daily activities is essential for a variety of applications, such as just-in-time content delivery or quantified self-tracking. Towards this, researchers often use customized wearable motion sensors tailored to recognize a small set of handpicked activities in controlled environments. In this paper, we design and engineer a scalable, daily activity recognition framework, by leveraging two widely adopted commercial devices: Android smartphone and Pebble smartwatch. Deploying our system outside the laboratory, we collected a total of more than 72 days of data from 12 user study participants. We systematically show the usefulness of time, location, and wrist-based motion for automatically recognizing 10 standardized activities, as specified by the American Time Use Survey taxonomy. Overall, we achieve a recognition accuracy of 76.28% for personalized models and 69.80% for generic, interpersonal models.

Keywords

Activity Routine Recognition; Wearable Sensors; Web Repository Exploitation; Crowd-Sensing Platform

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavior Sciences; I.2.6 [**Artificial Intelligence**]: Learning

1. INTRODUCTION

The most prominent electronic device we carry with us today is the mobile phone. Through its ubiquity, a multitude of onboard sensors, powerful processing units, and communication capabilities, it has been suggested as an ideal platform for context-aware, activity recognition [2] and large-scale behaviour monitoring [3].

BODYNETS 2014, September 29-October 01, London, Great Britain

DOI 10.4108/icst.bodynets.2014.257014

The context recognition community has investigated numerous sets of target activities for automatic recognition, such as detecting transportation modes [12], location-specific routines (e.g. sleeping or working) [11] as well as simple activity primitives like sitting, standing, running, or walking [2]. Yet, researchers [15] emphasizes that we do not "wear" our smartphones at all times. For example, it is imaginable that we leave our phones within reach, but on the desk. As a result, our mobile phones are not always able to detect our physical movements. Moreover, as the population carries it at a single location in the pocket [9], the variety of activities we can detect with onboard motion sensors is limited.

Recently, with the miniaturization of processing units and sensors, watch-like systems with similar capabilities as the smartphone are now facing the market. Contrary with the mobile phone, these devices are typically fixed in placement and constantly worn by the user. Paired with open platforms and software development kits, the distribution of applications for these wearable systems is easy and quickly scalable to large populations of users. As such, a new opportunity for activity recognition based on crowd data seems at hand. But the question that still remains: How useful is this ecosystem of devices for enhancing activity recognition systems? Moreover, which modalities are more accessible and useful for recognizing the activities of people in day-to-day life scenarios?

It has been often discussed in the community that time and location are good estimators for routine activities. Indeed, attempts of leveraging population-scale time use statistics [7, 20] have shown strong improvements in recognizing routinely executed activities when combined with wearable motion sensors. However, various activities, such as household cleaning or socializing, is not as easily pinpointed by aggregated time-use data. Similarly, location suggests certain activities. But since the location is often depicted with absolute geographic coordinates, it has been only considered for person-dependent activity recognition systems [11], inhibiting inter-person generaliziation. Furthermore, gestural motion captured from wrist worn-sensors has been used and shown to generalize well across users [5, 21], however, only in laboratory settings for fine grained activities.

In this work, we investigate these single modalities and their fusion for high-level daily activity recognition. We leverage the recently popularized smart-watch, Pebble[1], and the ubiquitously available

[1] `https://getpebble.com`

smartphone, for instrumentation. Specifically, we make the following contributions:

1. Using widely adopted commercial devices, we describe our design and implementation of a complete pipeline towards instrumenting and recognizing activities in natural day-to-day scenarios. Specifically, we use Pebble smart-watches and Android phones to upload sensor data and activity labels for server-side processing.

2. To evaluate our system, we conduct a user study involving 12 participants, where on average, each user marked approximately 87 activities spanning 157 hours. Aside from offering initial instructions for setup and usage, participants received no further guidance. This ensures realistic simulation for the usage of our system by a general crowd.

3. From the data collected, we model a comprehensive set of 10 daily activities as defined by the American Time Use Survey [18]. Analyzing our results, we provide thorough, per-class analysis of the usefulness of individual and fused modalities for classifying different types of activities.

2. RELATED WORK

Smartphone-based sensors offers naturalistic and population-scale sensing. A recent survey by Lane et. al [10] discusses current work and the promise of this emerging paradigm for sensing. However, they also point out the critical phone placement problem, where they emphasize it is difficult to predict how a user will "wear" their phone or even whether the phone will be near the user. Inline with this, Patel et. al. [15] reveals through an empirical study that the phone is only within a user's reach 58% of the time, on average. In our work, we alleviate this sensor placement problem by leveraging the sensing ability of the Pebble smartwatch, which is naturally placed on the wrist and waterproof for wearability throughout the day.

On the data processing side of scaling up the instrumentation of the crowd, researchers have examined technical frameworks for gathering population-scale physical sensor data as well as social media content for correlated analysis [19, 17]. In our work, we similarly offer scalable web-based data collection, in addition to data processing pipelines to complete the recognition chain.

Outside the laboratory, location and time have been shown as important cues for activity routine recognition. Early work by Liao et. al [12] used raw GPS traces with time to determine high-level routines of users, such as "sleeping" or "working". CenceMe [13] is another project that captures the activity of the user (e.g. "walking", "sitting") in addition to their dispositions for sharing on social media platforms. We base our target activity routines on the American Time Use Survey (ATUS) taxonomy [18], providing a formal and comprehensive basis for modelling relevant daily activities.

Towards leveraging population-scale data for comprehensive daily activity recognition, Partridge and Golle [14] first introduced the use of results from ATUS for activity recognition. Very recently, Borazio and Van Laerhoven have investigated the use of population-scale time use data to augment wearable sensor signals [6, 7]. In our paper, we extend these works by adding in nearby location semantics and wrist-based acceleration data to fuse signals across location, time, and motion.

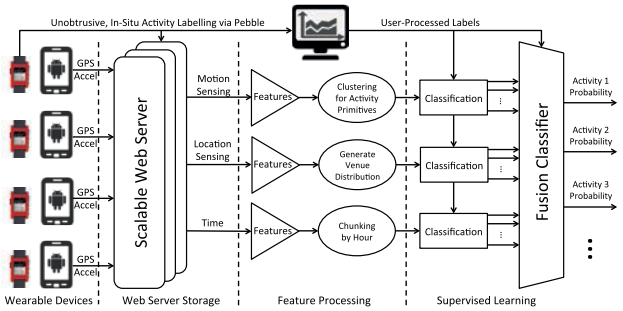

Figure 1: System architecture illustrating the four stages of data collection and processing.

3. SENSING AND LABELLING DAILY ACTIVITIES

In this section, we provide an overview and motivation for our sensing architecture. In the following section, we explain the specifics of the feature processing and model learning stages. A graphical illustration of the entire system is shown in Figure 1.

3.1 On-Body, Local, and Time

Popular wearable electronics are nowadays equipped with a rich set of sensors providing signals relevant to the user's activities. As mentioned, the Pebble is used for on-body sensing and activity labelling. We configure the Android-based smartphone app, Funf Journal [1], to collect and upload additional sensor readings onboard the smartphone, such as GPS.

To sense on-body motion, we leverage the Pebble smart watch's open platform for watchapp development and record wrist-based acceleration data. Onboard the Pebble watch, we calculate statistical features (mean and variance) and transmit this for logging on the smartphone.

To obtain the features describing the local vicinity of the user, we periodically poll the GPS signal of the smartphone. Instead of working with the raw GPS coordinates, we reverse-geocode the coordinates via Foursquare's venue lookup API to construct a vector of nearby venue types (e.g. [2 restaurants, 1 drugstore, 1 train station]). Examples of venue categories include: "Mexican Restaurant", "Piano Bar", "Aquarium", etc. The Foursquare Venue categorization system[2] contains 267 venue types, which are semantically arranged in a hierarchical fashion. The assignment of venue categories to venues is crowdsourced to Foursquare users. For our location features, we use the bottom-most layer of venue category to obtain the most fine-grained venue semantics. The purpose of this step to convert GPS coordinates to semantically meaningful venue distributions is to produce more generalizable features for learning interpersonal models as well as smoothing out the noise associated with possibly imprecise GPS positioning. To ensure the privacy and battery life of the user's mobile phone, the GPS sensor can be manually switched off by the user at any time during the study.

Since our sensor data and activity labels are timestamped for synchronization, we also calculate a time-base feature for no additional sensing cost. We do so by chunking the timestamp of signals by

[2] `https://developer.foursquare.com/categorytree`

Figure 2: An illustrations of the screens that a user sees when labelling activity routines via the Pebble interface. On the initial screen, the start and stop buttons are shown. Upon pressing the start button via the selectors on the right of the watch, a selection screen is displayed to mark the current activity routine. The user receives a confirmation message after confirming the selection of an activity (third photo). Finally, the current marked activity is shown after a few minutes into the activity.

hour, which is represented as a binarized 24-dimension indicator feature.

3.2 Activity Labelling

To minimize the labelling effort of users, we implement a watch application for the Pebble to allow unobtrusive, in-situ activity labelling. We show sample screens in Figure 2 and explain the simple steps of marking an activity in the caption below. To comprehensively capture various activities in day-to-day scenarios, we ask users to label tier-1 activities from ATUS taxonomy. Examples of these labels include "Eating or Drinking", "Consumer Purchases", or "Work-Related". In the ATUS taxonomy, these activities break down further into more precise activities (e.g. "Online Shopping" under "Consumer Purchases"), up to tier-3. However, to keep the labelling experience simple for this study, we only ask for tier-1 activity labels.

Compared to labelling on the smartphone, this is a seamless experience as the user would only need to press a few buttons on his/her wrist, therefore eliminating the need to access the phone, which may not always be within reach. In comparison with a diary-based approach where the user recollects a day's activity routines at the end of the day, our approach reduces labelling noise as the user labels while conducting an activity as opposed to having to recall the precise time duration afterwards.

To further enhance labelling accuracy, we build a web-based timeline visualization for the user to view, edit, or add activity annotations. This feature helps eliminate inevitable human labelling errors, such as forgetting to mark the stop of an activity routine.

4. FEATURE PROCESSING AND CLASSIFIER FUSION

From Figure 1, we depict three pipelines of processing followed by a fusion classifier. For all classification modules, we use the Random Forest classifier [8] with default parameters from [16]. Indeed, this is switchable for other classifiers capable of multi-class classification with probabilistic outputs.

Motion: The accelerometer data are treated in a similar way as in [4]. For the i-th 1-second data window, a vector $\mathbf{a}_i$ is built with the means and standard deviations of the three Pebble accelerometer axes, i.e. $\mathbf{a}_i = [\mu_x, \sigma_x, \mu_y, \sigma_y, \mu_z, \sigma_z]$. In the training phase, all the vectors $\mathbf{a}_i$ are clustered using Mini-Batch K-Means (K set empirically to 50) [16] and the corresponding K cluster centroids are stored. For each vector $\mathbf{a}_i$, a cluster distance vector ϵ_i of length K is obtained by computing the Euclidean distances between $\mathbf{a}_i$ and each of the K cluster centroids. Finally, a cluster similarity vector ξ_i is obtained from the cluster distance vector. Each element $\xi(k)$ of the similarity vector is computed from each element $\epsilon_i(k)$ of the cluster distance vector with the transformation: $\xi(k) = \exp\left(-\frac{\epsilon_i(k)}{\sigma_\epsilon}\right)$, where σ_ϵ is the standard deviation of the vector ϵ_i.

For each labelled training instance of duration D, a feature vector is obtained by averaging the D cluster similarity vectors calculated for each 1-second window belonging to the training instance. A classifier is trained with the feature vectors and the corresponding labels.

In the deployment phase, the cluster similarity vectors are again extracted from the accelerometer data, using the cluster centroids stored in the training phase. For each window of duration D, the corresponding feature vector is calculated and classified according to the trained model, outputting probability estimates for each activity class.

Location: Using raw GPS coordinates of the mobile phone, we query the Foursquare Venues API endpoint[3] to obtain a listing of venue in the vicinity of the GPS coordinate.

We process this list by counting the categories of venues to obtain a venue semantic distribution vector $[v_1, v_2, ..., v_T]$ where T is the total number of venue categories observed. As we typically have multiple GPS samples within a window instance, we normalize the venue semantics vector by the number of GPS samples. Essentially, the feature space for the location classifier is a sparse count matrix of venues types, which takes a dimension of $N \times T$ for N instances. Again, from the trained classifier, we output the classification probability of each activity.

Time: Given the duration of each activity window, we convert the unix timestamp to local time and bin the duration of the window into 24 hourly bins for each day of the week. Then, for each instance, we derive a binarized feature vector $[t_1, t_2, ...t_{168}]$ to indicate the hour(s) of the activity window. Similar to the other modalities, we train a classifier on this feature space and output the classification probabilities.

Fusing Classifier: Our system utilizes a stacked classifier approach to fuse the individual modalities at the classifier level. Therefore, we concatenate the probability outputs of the aforementioned classifiers to create a feature matrix of $M \times C$, where M is the number of modalities and C is the number of classes. As the Random Forest classifier naturally handles multi-class classification, we directly learn a mapping between the single-modal classifiers' per-class probability outputs and the ground truth label.

Also known as late fusion, classifier-level fusing is practical when not all modalities may output data for the same instance. In this

[3] `https://developer.foursquare.com/docs/venues/search`

Activity Routine	Number of Marked Activities	Duration (hours)	Instances Derived
Working or Work-Related	266	499.18	3474
Travelling	189	124.01	487
Eating/Drinking	166	120.89	677
Personal Care	126	640.32	4407
Socializing, Relaxing, and Leisure	83	162.37	980
Household Activities	45	21.92	68
Education	44	105.44	416
Sports, Exercise, and Recreation	23	52.53	201
Consumer Purchases	13	5.19	16
Receiving Services	1	0.47	2

Table 1: Summary statistics on the quantity of activity data collected from 12 study participants.

case, the fused classifier is only based on modalities that contained data. For comparison purposes, we also implement feature-level fusion, where feature sets from different modalities are scaled and concatenated to form a fused feature space. The classifier is then trained on the concatenated features. Unlike classifier-level fusion, this approach would suffer in performance if not all modalities are available. For such instances, we insert zeros for the missing features.

5. SYSTEM DEPLOYMENT AND DATA COLLECTION

5.1 User Study Deployment

We deployed our activity logging system to 12 users over the course of a month. Participants logged a median of 11.5 days while the maximum and minimum number of days logged were 6 and 18, respectively. Our user-base consists of staff and student members from a university environment.

In our user study, we provide basic instructions on how to utilize our activity logging Pebble watchapp. We ask users to occasionally record activity routines as they conduct them. Therefore, a start marker would be entered once a user start an activity routine and an end marker is to be recorded once the activity routine is finished. We also instruct the users that they can long press the stop button on the Pebble smartwatch if they forget to enter a stop marker in a timely fashion. Since we are interested in activity routines that typically span multiple hours of the day, we allow users to disregard short activities or periods where they are transitioning from one activity to another (e.g. brief walk of a few minutes from class to the cafeteria). At the end of the study, we display an interactive timeline through a web interface, allowing users to review and correct their activity timelines. Activity routines that are marked with an untimely stop marker (via long press) were highlighted in red on the web interface. For users who do not prefer to use their own phone, we provide either a Samsung Galaxy S4 or S3 mini for use as the user's primary mobile phone.

Aside from helping the users install and setup the logging applications, we provide no further guidance for the duration of the study aside from the instructions mentioned above. Our intention is to deploy this user study in the most naturalistic manner possible as if this system was used by a general crowd.

5.2 Dataset Description

Although the original ATUS taxonomy contains 18 tier-1 activity categories, in our study, we only include the activities that received at least one label from our participants. In Table 1, we detail the quantity of data collected for the individual activities. In columns 2 and 3 of Table 1, we list both the number of activities marked for each category as well as the number of hours labelled from all participants. In Figure 3, we plot the distribution of activities from all users, summed hourly, over a 24-hour period. The activity durations are chunked by hour and the proportion of activity categories are plotted. From the plot, we can spot intuitive patterns over the day, for example: sleeping during the morning hours (a component of Personal Care) or work-related activities throughout the day although indented by eating and drinking around lunch time. It is also interesting to note the intuitive transition of activity proportions, e.g. from "Travelling" (to work) to "Work-Related" to "Travelling" (to home).

Preprocessing the dataset, we filter out activity labels less than 20 minutes in duration. We also filter out marked activities that contain no sensor data from either location or motion modalities. Missing GPS data may be caused by the user manually turning off the GPS sensor on their smartphone. Missing motion data from the Pebble watch is caused by the user being away from their phone for an extended period of time, such that the local data storage on the Pebble is filled over its capacity.

In our study, approximately 18% of the marked activities are discarded due to missing data, including one user who was filtered out completely due to sparsity of labels. In the remaining data from 11 users, we create 50% overlapping windows of 20 minutes each as instances for our machine learning model. In the third column of Table 1, we indicate the number of instances derived for each activity category after applying the filtering and windowing steps.

6. SYSTEM EVALUATION

In this section, we evaluate our system's ability to recognize daily activity routines using individual modalities as well as fused modalities. For personalized recognition models, we cross-validated using the leave-one-day-out scheme. We also investigate our system's ability to generalize for interpersonal models, where we conduct leave-one-person-out cross-validation. We start our analysis by examining the overall performance of our system through classification accuracy. Then, we provide finer-grain analysis of system performance for specific activities that we target.

6.1 Overall Accuracy

In Table 2, we list the testing accuracy achieved with the personal and interpersonal models when using individual and fused modalities. Comparing the feature sets used, we bold and underline the top performing feature set and only bold the second place performance.

Looking down each column of Table 2, we see early fusion (at the feature level) or late fusion (at the classifier level) provide top performance for most users. Unsurprisingly, fusing the three modalities typically results in higher accuracy, especially when individual modalities are not performing well (e.g. user 9). Across users, classifier level fusion delivers only slightly better performance results than feature level fusion. However, the real benefit of late fusion lies in its tolerance of missing data. Practically, this is an important consideration as signals from different modalities may not be available simultaneously at all times.

Comparing the accuracy performance of personalized and interpersonal models, we can immediately notice the use of acceleration signal from the wrist is not very generalizable between users. Therefore, personalized training data should be used in this modal-

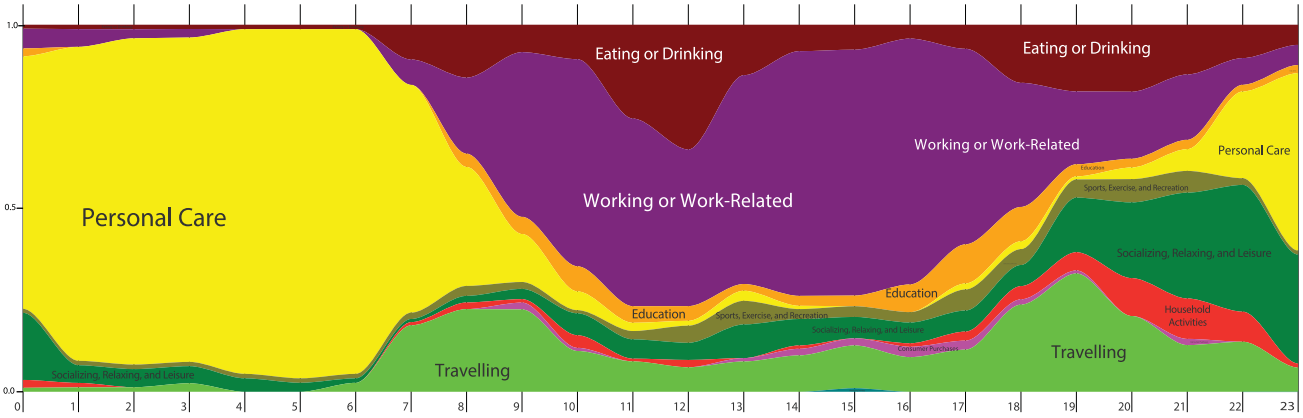

Figure 3: Timeline of accumulated activity labels from all users over a merged 24-hour period. Binning activity periods by hour, counts of activities are normalized by the total number of activity counts in each hourly bin. The layers from top to bottom are: Eating or Drinking (maroon), Working or Work-Related (purple), Education (orange), Personal Care (yellow), Sports, Exercise, and Recreation (olive), Socializing, Relaxing, and Leisure (green), Household Activities (red), Consumer Purchases (fuchsia), Travelling (lime), and Receiving Services (teal).

	Motion	Location	Time	Feature Fusion	Classifier Fusion
User 1	0.26	***0.42***	0.14	**0.41**	0.35
User 2	0.75	**0.75**	0.51	**0.81**	0.74
User 3	0.83	0.87	0.68	**0.90**	**0.88**
User 4	0.61	0.71	**0.78**	0.70	**0.76**
User 5	0.66	**0.75**	0.60	**0.75**	0.73
User 6	0.68	0.72	0.69	**0.77**	**0.73**
User 7	0.71	0.40	**0.86**	0.66	**0.88**
User 8	0.82	**0.84**	0.82	**0.85**	0.79
User 9	0.36	0.32	0.35	**0.46**	**0.41**
User 10	0.78	**0.99**	0.75	**0.95**	0.94
User 11	0.70	0.73	**0.81**	0.71	**0.78**
Median Accuracy	0.70	0.73	0.69	**0.75**	**0.76**
Interpersonal Accuracy	0.53	0.54	**0.69**	**0.66**	0.65

Table 2: Testing accuracy of different modalities for personalized and interpersonal classification models. For personalized model, we show the median of the average testing accuracy over the 11 results.

ity. We also notice a large drop in performance when using location-based features. We believe there are two explanations as to why the venue-based interpersonal model suffers in performance. First, aside from some participants who work in the same vicinity, all participants live in different locations across the city. Therefore, the majority of participants have very different venue distributions in their home vicinity, where a large proportion of time is spent. Second, even though we believe semantic location features (e.g. restaurants, libraries) offer some generalization benefits over raw GPS location, many residential neighbourhoods contain few labelled Foursquare venues, unlike city-centres. As a result, we believe location-based features are more useful for activities typically conducted in non-personal locations, such as "Work-Related".

On the other hand, the top performing feature set for interpersonal prediction is time, which achieves very similar performance as personalized models. Intuitively, this shows different people are not only routine according to their own schedules, but also to the schedules of others. Again, this is not surprising as most people tend to work 8-6 during weekdays and sleep during the hours of early morning. Nonetheless, our results supports the use of aggregated daily rhythms to identify long-spanning activities, such as investigated in the work of [7, 14, 6].

Although the accuracy performance mentioned is reasonable for 10 classes of activity routines, we note the skewed distribution of the classes, where work-related activities and personal care (including sleep) dominate. In the following analysis, we take a more detailed look at the performance of recognizing individual activities.

6.2 Recognition Performance of Individual Activities

In Figure 4, we illustrate the distribution of mean precision and recall scores of personalized models. We use the standard plotting definition of box-plots and show the boundaries of the interquartile range (or midspread) of the data with the box boundaries, while the median is plotted as a horizontal line inside the box. The whiskers bound data within 1.5 times the interquartile range. Outside of this range, the stars indicate outlier scores of select participants.

Immediately, it is clear that the activities with a larger number of instances (e.g. "Personal Care" and "Work-Related") are highly distinguishable while some less frequently labelled activities (e.g. receiving services with 2 instances or consumer purchases with 16) are not distinguishable at all.

We notice that different individual modalities are useful for certain activities. For example, it is not surprising wrist-based accelerometer features are the most useful for distinguishing sports-related activities while time is the best individual modality to distinguish attending fixed-schedule lectures (i.e. "Education"). Just as intuitive is "Work-Related" activities bound to specific locations, thereby allowing location-based features to deliver the highest performance. It is interesting to note that accelerometer data is quite useful for

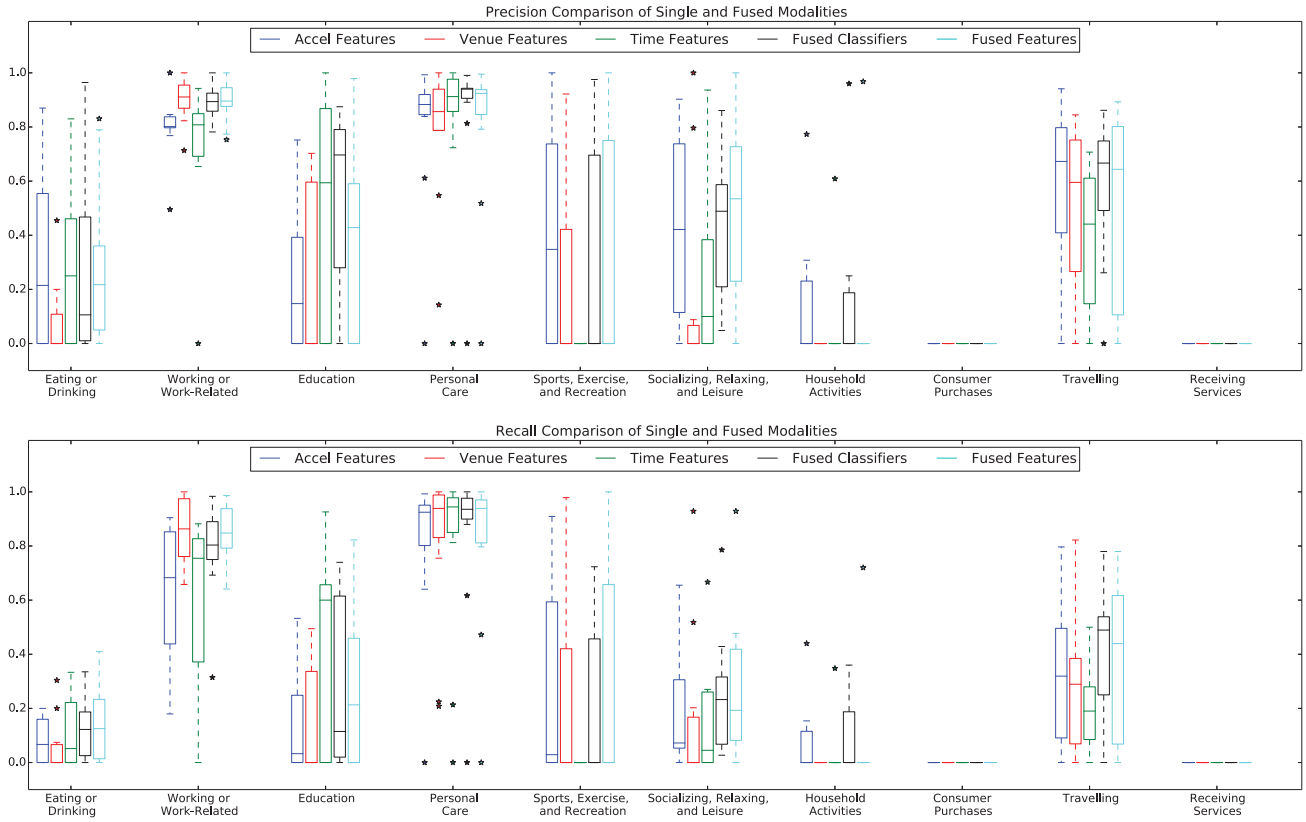

Figure 4: We plot the distribution of average precision and recall scores for 11 participants. The box-plots are used to simultaneously illustrate where the majority of the data lies as well as where outliers are.

distinguishing activities not necessarily bound to locations or time, such as "Socializing, Relaxing, and Leisure" or "Travelling".

In comparison, we observe modal fusion (especially late fusion) approximately maintains or improves median performance over most activities. The only exception we see is in "Education", where time outperforms all other classifiers due to the routine nature of lectures. Although not always the top performer, late fusion consistently delivers near-top performance compared with other modalities. This motivates its use for activity recognition systems aimed at detecting a comprehensive set of daily activities.

7. LIMITATIONS AND FUTURE WORK

In this work, we make a first step towards comprehensive daily activity recognition with widely adopted commercial devices. Below, we highlight some limitations and future work to address them.

Large-Scale System Deployment: Although our system is designed to scale, we have only shown a proof-of-concept through the small-scale user study illustrated here. One key challenge for us is to engineer the necessary incentivization strategies to encourage crowd contribution of activity data and labels. One possible solution we are exploring is to extend the activity-reviewing web interface. One can imagine this implemented as a life-logging portal, where service such as activity tracking and time-use analysis can provide value to the user.

Feature Engineering: Contrary to our original expectations, we did not notice a high degree of interperson generalizability in our location-based features, especially compared to simple time-based features. Intuitively, the existence of "nearby" venue types is indicative of activities (e.g. "Education" at the lecture hall), however, there is also a significant amount of noise as many venues are multi-purposed (e.g. one's home). In the future, we intend to examine the usefulness of richer location descriptions, such as location-specific, crowd-generate reviews or tips, to enrich the location modality.

Multi-Modal Fusion: Although we noticed improvement gains using feature-level and classifier-level fusion, the performance gain was only incremental compared to the best performing single modality. Although outside the scope of this work, we intend to investigate more sophisticated multi-modal fusion strategies in future work.

8. CONCLUSION

We presented and investigated a commercially available ecosystem of mobile and wearable devices for daily activity recognition. Our system was successfully deployed as a proof-of-concept user study, where we tested naturalistic usage. We collected more than 72 days of data in total from 12 users and trained personalized and interpersonal classification models based on three modalities: motion, location, and time.

With personalized models, classifier-level fusion achieved the high-

est median accuracy with 76.28%. Upon deeper examination of classification performance across individual activities, we found that fusion typically outperforms single-modality classification in terms of precision and recall. Given the variability of single-modality performance across different activities, we believe classifier-level fusion should be used for comprehensive daily activity recognition. Evaluating generalization capability of interpersonal models, we found that the highest testing accuracy was achieved with time-based features, at 69.80% while classifier-level fusion achieved 65.46%. Understandably, time plays a key role in identifying inter-person synchronized activities (e.g. sleeping). However, we find that the fused classifier is able to make gains for activities not necessarily time-synchronized between people, such as "Education", and "Sports-Related". Again, we believe it is beneficial to leverage multiple modalities if a wide range of activities is to be recognized.

9. ACKNOWLEDGEMENTS

We would like to thank Yu Jiang, Masafumi Suzuki, and Stephanie He for assisting with implementation efforts. We thank Rahul Bhagat at Pebble for supplying us with Pebble watches. We are also grateful to the reviewers for their comments and to the participants of our user study. This work is partially supported by the Hasler Foundation through the SmartDAYS project.

10. REFERENCES

[1] N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland. Social fmri: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7(6):643–659, 2011.

[2] M. Berchtold, M. Budde, D. Gordon, H. R. Schmidtke, and M. Beigl. Actiserv: Activity recognition service for mobile phones. In *Wearable Computers (ISWC), 2010 International Symposium on*, pages 1–8. IEEE, 2010.

[3] U. Blanke, T. Franke, G. Tröster, and P. Lukowicz. Capturing crowd dynamics at large scale events using participatory gps-localization. In *The 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 2014.

[4] U. Blanke and B. Schiele. Daily routine recognition through activity spotting. In *International Symposium on Location and Context Awareness (LoCA)*, May 2009.

[5] U. Blanke, B. Schiele, M. Kreil, P. Lukowicz, B. Sick, and T. Gruber. All for one or one for all? – combining heterogeneous features for activity spotting. In *7th IEEE PerCom Workshop on Context Modeling and Reasoning (CoMoRea)*, Mannheim, Germany, 2010.

[6] M. Borazio and K. Van Laerhoven. Improving activity recognition without sensor data: a comparison study of time use surveys. In *Proceedings of the 4th Augmented Human International Conference*, 2013.

[7] M. Borazio and K. Van Laerhoven. Using time use with mobile sensor data: a road to practical mobile activity recognition? In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, page 20. ACM, 2013.

[8] L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, Oct. 2001.

[9] F. Ichikawa, J. Chipchase, and R. Grignani. Where's the phone? A study of Mobile Phone Location in Public Spaces. In *2nd International Conference on Mobile Technology, Applications and Systems*, pages 1–8, 2005.

[10] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

[11] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition. In *Neural Information Processing Systems (NIPS)*, 2005.

[12] L. Liao, D. J. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. *Artificial Intelligence*, 171(5):311–331, 2007.

[13] E. Miluzzo, N. D. Lane, S. B. Eisenman, and A. T. Campbell. Cenceme: Injecting sensing presence into social networking applications. In *Proceedings of the 2Nd European Conference on Smart Sensing and Context*, EuroSSC'07, pages 1–28, Berlin, Heidelberg, 2007. Springer-Verlag.

[14] K. Partridge and P. Golle. On using existing time-use study data for ubiquitous computing applications. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 144–153, New York, NY, USA, 2008. ACM.

[15] S. N. Patel, J. A. Kientz, G. R. Hayes, S. Bhat, and G. D. Abowd. Farther than you may think: An empirical investigation of the proximity of users to their mobile phones. In *UbiComp 2006: Ubiquitous Computing*, pages 123–140. Springer, 2006.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[17] T. Phan, S. Kalasapur, and A. Kunjithapatham. Sensor fusion of physical and social data using web socialsense on smartphone mobile browsers. In *Proceedings of the 11th Annual IEEE Consumer Communications and Networking Conferece*, 2014.

[18] K. J. Shelley. Developing the american time use survey activity classification system. *Monthly Lab. Rev.*, 128:3, 2005.

[19] W. Sherchan, P. P. Jayaraman, S. Krishnaswamy, A. Zaslavsky, S. Loke, and A. Sinha. Using on-the-move mining for mobile crowdsensing. In *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on*, pages 115–124. IEEE, 2012.

[20] K. Van Laerhoven, D. Kilian, and B. Schiele. Using rhythm awareness in long-term activity recognition. In *Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on*, pages 63–66. IEEE, 2008.

[21] A. Zinnen, U. Blanke, and B. Schiele. An analysis of sensor-oriented vs. model-based activity recognition. In *Proceedings of the 13th IEEE International Symposium on Wearable Computers (ISWC)*, 2009.