Semantic Activity Classification Using Locomotive Signatures from Mobile Phones

Zhixian Yan[1], Dipanjan Chakraborty[2], Archan Misra[3], Hoyoung Jeung[4], Karl Aberer[1]

[1]EPFL Switzerland {zhixian.yan, karl.aberer}@epfl.ch

[2]IBM Research Lab India cdipanjan@in.ibm.com

[3]Singapore Management University Singapore archanm@smu.edu.sg

[4]SAP Research Australia hoyoung.jeung@sap.com

***Abstract*—We explore the use of mobile phone-generated sensor feeds to determine the high-level (i.e., at the semantic level), indoor, lifestyle activities of individuals, such as cooking & dining at home and working & having lunch at the workplace. We propose and evaluate a 2-$Tier$ activity extraction framework (called SAMMPLE[1]) where features of the low-level accelerometer data are first used to identify individual locomotive micro-activities (e.g., sitting or standing), and the micro-activity sequence is subsequently used to identify the discriminatory characteristics of individual semantic activities. Using 152 days of real-life behavioral traces from users, our approach achieves an average accuracy of 77.14%, an improvement of 16.37% from the traditional 1-$Tier$ approach, which directly uses statistical features of the accelerometer stream, towards such activity classification tasks.**

***Keywords*-activity recognition, semantic activities, sensor data analytics, context mining, pattern mining**

I. Introduction

Research on "people sensing" [3] primarily focuses on using smartphone-embedded sensors (e.g., GPS, accelerometers, gyros) to infer either an individual's movement behaviors [4][21] or specific micro-activities of a person, such as sitting, walking, running, or cycling [1][2][16]. In this context, *micro-activity* is referred to as a specific set of these locomotion or postural states. A longer-term research goal is, however, to combine such micro-activity sensing with a better awareness of environmental context to infer higher-level *macro* or *semantic* activities—e.g., using the movement and interaction pattern of a group of individuals in a home to deduce that they are "having dinner" or "taking a smoke break", as opposed to simply 'sitting' or 'standing'. We call them semantic activities[2]. The ability to infer such activity context at the semantic level will greatly enhance the a acceptability of many emerging applications, in areas such as context-aware notification [12] and healthcare [9].

Such semantic-level activity mining has traditionally been investigated in highly-instrumented *smart home* environments, using object-embedded and multiple wearable sensors [8][11][18], or for outdoor activities by combining geographic data with real-life GPS traces [21][24][25]. Such approaches are generally inapplicable for lifestyle activities performed in *un-instrumented* indoor spaces, both personal (e.g., home, office) and commercial (e.g., shopping malls, movie theaters).

Accordingly, this paper addresses the problem of inferring an individual's **indoor semantic activities** based on smartphone-generated sensor traces in 'out-of-the-lab' environments (an objective receiving increasing emphasis [14]). More specifically, we aim to infer the semantic activities based on an individual's fine-grained locomotive or postural behavior, captured solely via a phone-embedded accelerometer sensor. We focus on the accelerometer as it constitutes the most commonly-available, easily programmable, low-energy sensor in current personal mobile devices.

Research Questions: Our quest to use accelerometer-generated data as a basis for semantic activity detection raises two key research questions:

a) In real life, do an individual's semantic activities possess enough regularity and discriminatory power, in terms of accelerometer-based features, to permit unambiguous classification of an unlabeled semantic activity? And,
b) If so, how do we design an activity classification framework that identifies and leverages upon such discriminatory features, and what level of accuracy does it achieve for different types of daily-lifestyle activities?

Key Contributions: To address these questions, this paper uses real-life observational traces to make the following key contributions:

1) We present a novel 2-$Tier$ process of semantic activity inferencing that first transforms the raw accelerometer data into a sequence of an individual's micro-activities, and then employs statistical feature extraction & mining on the micro-activities to identify the most likely semantic activity. More specifically, the accelerometer readings, associated with an unlabeled semantic activity, are first processed to derive a sequence of micro-activities while GPS readings are used to identify whether the individual is at home or office. The studies in this paper are restricted to these 2 semantic locations, where a typical user spends the majority of her time. This location-tagged micro-activity sequence is then mined for features to extract the most likely semantic activity, e.g., cooking at home, dining at home, coffee break at office.

This 2-$Tier$ activity classification framework is unique, and distinct from prior work (e.g., [13][20]) that classifies activities directly using statistical features computed from the raw sensor

[1] SAMMPLE: Semantic Activity Mining via Mobile Phone-based Locomotive Estimation

[2] "Semantic activity" and "High-level activity" are used synonymously in the paper

streams. We demonstrate that the 2-$Tier$ process outperforms existing approaches, as it *a)* is more robust to underlying sensor noise, *b)* can better accommodate the day-to-day behavioral variations in semantic activities and *c)* provides a more intuitive, locomotion-based view for comparing different semantic activities.

2) We define and evaluate two discriminatory feature extraction techniques operating on the intermediate, micro-activity sequence, in order to classify specific semantic activities. The first approach analyzes only the total duration of different types of underlying micro-activities, resulting in a classification accuracy of $\sim$60-80%. The second approach additionally considers the temporal order of these micro-activities, improving the classification accuracy by 4-15%. In the end, using discriminative features from micro-activity sequences helps SAMMPLE to gain 16.37% accuracy improvement from the traditional 1-$Tier$ method.

3) In contrast to laboratory studies where the placement and orientation of accelerometers/phones are constrained to specific on-body positions, we focus on applying this process in a *naturalistic environment*. Hence, to ensure that our results are valid under naturalistic conditions, this paper utilizes two different user-generated data traces from 5 users. The first data set (**MICRO-SHORT**) is used to determine the best features for classifying micro-activities in controlled but naturalistic conditions, where the smartphone's usage and on-body position varied dynamically. The second data set (**SEMANTIC-LONG**) captured accelerometer readings from the phones as users went about performing their daily lifestyle-based semantic activities, for a period of 8 weeks.

Our results on achievable accuracies for semantic activities, using a *single* phone-embedded accelerometer under naturalized usage, is a first-of-a-kind study, and provide valuable insights to the "smartphone-based sensing" research community.

II. Related Work

Activity mining is an active research area spanning different domains such as web log mining, mobility data mining, and recently, social network data mining. We cover the work closest to our focus in this paper.

Locomotion Learning from Accelerometer Feeds: Prior work (e.g., [1]) largely focused on feature extraction and classification of representative locomotions (e.g., walking, running, cycling) using data from *multiple* body-worn accelerometers under "lab" environments. Recent approaches [8][16] focus on accurate prediction of an individual's locomotive state using a single accelerometer, but fixed to a pre-defined body position. [14] addressed the problem of locomotion and posture prediction (i.e., micro-activity classification) using cellphone-embedded accelerometers, but for reasonably well-separable activities. While our study aligns with this latter direction, we additionally focus on less-separable micro-activities, and on understanding how various feature choices perform under naturalistic phone usage, with the phone's on-body location & orientation being subject to dynamic changes.

Semantic Activity Mining and Smart Environments: Several research prototypes (e.g., [4][21]) have focused on outdoor semantic activity inferencing and annotation, where GPS-based movement history of an individual is combined with GIS-assisted location semantics (e.g., shopping mall) to infer the person's semantic activity. In smart homes literature, several systems (e.g., the MIT PlaceLab [18]) have studied the inference of semantic activities, using on-body and on-object sensors (e.g., RFID) to capture the interaction of inhabitants with individual objects & appliances (e.g., doors, refrigerators). Broadly speaking, these approaches either focus on outdoor activities or are *infrastructure-dependent*. In contrast, we focus on the challenge of inferring semantic activities in un-instrumented indoor spaces, using a single daily phone-based accelerometer.

Activity Recognition from Mobile Phones: Approaches for mobile phone-based activity recognition principally aim to leverage upon multiple phone-embedded sensors to recognize the user's context [7][14][15]. For example, [6] demonstrated that accelerometers and microphones provided good features for activity recognition (e.g., walk, run, talk, cook, eat) while [14] developed an 'on-phone' classification system to detect events using multiple phone sensors. The primary focus is to discover how a combination of phone sensors helps to improve the accuracy of activity detection. It is natural that this accuracy should increase with the use of additional sensors. We however investigate a complementary question: *To what extent can the accelerometer sensor alone be used discriminate indoor semantic activities at different locations?*

Hierarchical Activity Recognition: [19] studied an unsupervised learning of key sensor signatures to study higher-level activities (e.g., "having a snack" vs. "having a meal") from underlying labeled activities such as "moving from chair to fridge", which were derived from object interaction-based sensor readings. [10] applied a statistical topic model-based approach (using LDA - Latent Dirichlet Allocation) to perform unsupervised identification of key high-level activity routines and discover their associations with a large set of low-level activities, each corresponding to a topic. While sharing a similar goal, we have a couple of important differences: *a)* These approaches all utilize multiple sensors (multiple accelerometers or RFID/object interaction sensors) and can thus identify a much richer set of lower-level activities, as opposed to 'locomotive' activities detectable by only a single accelerometer; *b)* They focus on unsupervised discovery of the relationships between *a hierarchy of semantic activities*. We focus on uncovering rich feature sets that best classify a low-level accelerometer stream to semantic activities, under natural lifestyle-based day-to-day variations.

III. The SAMMPLE Inference Approach

The SAMMPLE approach is a 2-Tier classification process that infers an individual's semantic activity, using *micro-activities* as an intermediate step. Fig. 1 illustrates the two-level hierarchy of SAMMPLE.

In the lower layer, the raw accelerometer data corresponding to a semantic activity is first partitioned into a sequence of non-overlapping "frames" of small duration, denoted by T_f–e.g., 5secs. We make a reasonable assumption that each

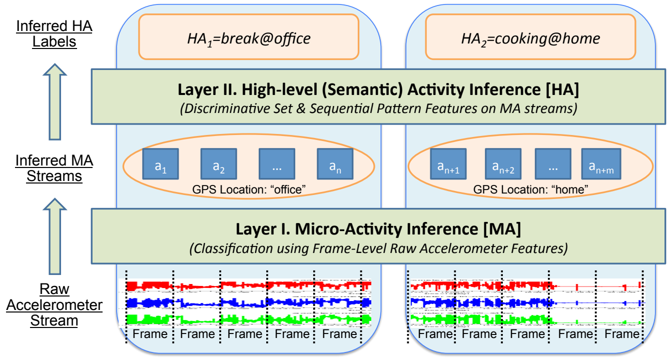

Fig. 1: Our 2-Tier Semantic Activity Inferencing Process

frame corresponds to a specific locomotive/postural state of the individual (i.e., micro-activity). We extract statistical features from each frame, and employ classification algorithms to map each frame into a corresponding micro-activity.

The upper layer of SAMMPLE then accepts this intermediate sequence of inferred 'micro-activities' and employs a *separate* classification procedure to label the unknown semantic activity, using appropriate *features defined on this micro-activity sequence*. Note that while each frame is of constant size, the duration of a semantic activity (such as "cooking") can vary, implying a corresponding variation in the size of the micro-activity sequence.

The 2-Tier approach is motivated by the belief that defining features in terms of this intermediate micro-activity sequence enables us to better accommodate the inherent natural variations in the *duration* and the *specific sequential way* of performing a semantic activity. For example, it is likely that different "cooking" instances, while of different duration, have a characteristic relationship among the underlying 'walk', 'stand' and 'sit' frames. For e.g., perhaps "cooking" implies that $\sim 60\%$ of the frames are 'stand'. Similarly "office_break" might have a temporal order between "standing", "loitering" and "sitting". These characteristics are almost impossible to discover and define in terms of typical low-level statistical features, such as the magnitude or FFT components of the raw accelerometer data. Note also that, in contrast to prior work [1][14] which defines a variety of accelerometer features for classifying micro-activities, we are among the first to address the question of *"what micro-activity features can help classify a semantic activity?"*

A realization of our SAMMPLE approach must address two key challenging issues:

- *What set of statistical features defined over raw accelerometer data provide high accuracy in classifying micro-activities under naturalistic conditions*? For the 2-$Tier$ framework to work well, micro-activities at the lower layer must be classified accurately, to yield a *bona fide* sequence of micro-activities for the upper layer. We have to take into account the reality that the same micro-activity can generate different accelerometer-based statistical signatures at different times, depending on variations in the phone's placement & orientation (e.g., placing it in one's trouser, laptop bag or hand).
- *How do we identify the set of discriminative features that help us to classify an unknown semantic activity, given a sequence of micro-activity labels?* Clearly, the learning has to be *personalized*, as individuals vary in their semantic activity patterns (e.g., how they cook).

We address these two questions in the following two sections. Though we experiment with only indoor semantic activities, our framework is not limited to these indoor locations. For the purposes of this paper, we assume that the accelerometer samples are accompanied by a GPS-based location tag that identifies whether the activity occurred at home or in the office, implicitly restricting the possible choices for the unknown semantic activity. Prior work [21] has shown how such semantic locations can be extracted from GPS data, and is incorporated in our framework.

SAMMPLE implicitly assumes, similar to earlier studies exploring rich feature sets [8][18], that the duration (i.e., start and end times) of a semantic activity is known. Ideally, the classification system should also detect when an activity transition occurs. Detecting such "change points" in continuous activity streams is a hard problem under investigation [23], and is out of scope of this paper. As such, while SAMMPLE can be applied to an un-delineated stream, its classification accuracy might suffer when activity transitions occur.

Notations and Problem Formulations: SAMMPLE's input consists of an accelerometer data segment $\mathcal{A}$, which is a sequence of periodically sampled[3] accelerometer records corresponding to an unlabeled semantic activity (HA) instance. Each record A_i includes acceleration (x, y, z), timestamp (t), and semantic location (sl). The output should be an inferred HA label from a set of distinct semantic activity labels $\mathcal{HA}$=$\{HA_1, \ldots, HA_H\}$. Therefore, the goal of SAMMPLE is then – *given a accelerometer data segment ($\mathcal{A}$) corresponding to an unknown HA of varying duration, find the* most-likely *HA label associated with this unknown HA instance*.

In the lower layer, SAMMPLE converts each HA instance data ($\mathcal{A}$) to a *Micro-Activity Sequence* $\mathcal{MS}$, where each element is an inferred MA label by using the raw data segment $\mathcal{A}$. The candidate MA labels are a set of M distinct micro-activities, i.e., $\mathcal{MA}$=$\{MA_1, \ldots, MA_M\}$. This paper studies 7 (M=7) micro-activities: {'sit', 'sit active', 'walk', 'loiter', 'bursty move', 'stand', 'using stairs'}. These were chosen based on users' feedback of micro-activities commonly associated with their daily lifestyles at home and office. While most MA labels are self-descriptive, the non-obvious ones are described in Table I. Therefore, the lower layer can be formulated as: MA_Inference($\mathcal{A}$) $\rightarrow \mathcal{MS}$. In the upper layer, SAMMPLE infers the unknown HA label for the MA sequence $\mathcal{MS}$, i.e., HA_Inference($\mathcal{MS}$) $\rightarrow HA_k$. Table II summarizes these important symbols used in the later sections.

TABLE I: Descriptions of some non-obvious Micro Activity Labels

Name of MA Label	*Exemplary Description of Activity*
sitActive	sitting but being active (e.g., shaking legs, stretching, ..)
sit	sitting in a static way
loiter	walk at a slow pace with stops, walk inside office rooms
burstyMove	jerky movements (e.g., get up from chair, movements inside kitchen)

[3]The results in this paper correspond to a sampling frequency of 30 Hz

TABLE II: Notations of symbols

Symbol & Definition	*Description*
$\mathcal{A}$=$\langle A_1, A_2, \ldots, A_n\rangle$	an accelerometer data segment
A_i=$(x_i, y_i, z_i, t_i, sl_i)$	a tuple including acceleration (x_i, y_i, z_i), timestamp t_i and associated semantic location $sl_i (\in \mathcal{SL})$
$\mathcal{SL}$ = {"*home*", "*office*"}	two restricted semantic locations in this paper
$\mathcal{HA}$=$\{HA_1, \ldots, HA_H\}$	H distinct semantic activity labels, e.g. *office_break*
$\mathcal{MA}$=$\{MA_1, \ldots, MA_M\}$	M distinct micro activity labels, e.g. *sit*
$\mathcal{MS}$	a sequence of MA labels generated from a HA instance in terms of a raw accelerometer segment $\mathcal{A}$
$\mathcal{MS}^{(k)}$	the MA sequence for a training (or testing) HA instance with known (or inferred) HA label, i.e., HA_k

IV. Lower Layer: Micro-Activity Inference

This section investigates a broad set of statistical features and classification algorithms that operate on the accelerometer data ($\mathcal{A}$), and empirically establish that good MA classification accuracy can be achieved under naturalistic conditions.

Broadly speaking, features evaluated in past work [14][1] can be grouped into two types: **1) Orientation-Dependent Features:** These features are computed separately on each of the three axes, x, y, z (e.g., mean values $\{\bar{x_i}, \bar{y_i}, \bar{z_i}\}$ of the raw data in frame F_i). In general, we can expect orientation-dependent features to be useful when the phone's orientation relative to the earth's horizontal remains constant. **2) Orientation-Independent Features:** These are either: (1) an orientation-insensitive combination of the readings from the axes (e.g., the variance of the acceleration magnitude $\sqrt{x^2 + y^2 + z^2}$); (2) associated with acceleration values *projected to a reference frame, e.g., the ground*. We expect orientation-independent features to be more robust to variations in a phone's orientation/on-body location, but offer lesser resolution than their orientation-dependent counterparts.

Past work has used *exclusively* either orientation dependent or orientation-independent features (e.g., [14]) and looks to distinguish well-separable micro-activities (e.g., sit, walk, jog, cycle). In contrast, we explore various *combinations* of these two feature classes, and evaluate their ability to classify the locomotive activities we encountered in SAMMPLE, which have varying degrees of similarity (e.g., 'walk' and 'loiter').

Feature Extraction: We consider a feature vector, consisting of both time and frequency domain features from: (1) the 3D axis of the phone, referred to as 3D-features; (2) A projection of the readings on the gravity direction ($\vec{p}$) and the plane perpendicular to gravity($\vec{h}$), which makes it orientation-independent (referred to as 2D-features). We use the fact that the mean of accelerometer readings, computed over a long time period gives an estimate of g [14], to project the raw signal to this "2D" reference frame. For the frequency domain, the features are computed by first transforming the (x_i, y_i, z_i) segment into a 250-point FFT vector [16]. Finally, a total of ~70 features are used per frame (F_i). Table III summarizes the feature types. To reduce random errors of sensor readings, a state-of-the-art calibration technique on the Nokia N95 [22] is used to calibrate the sensor readings, before feature extraction.

Classification Algorithms: We experimented with a wide variety of state-of-the-art classifiers. We note that our principal goal is not to devise any new feature or classification algorithm, but empirically determine features choices & classification algorithms that provide high MA classification accuracy in naturalized environments.

A. Description of MICRO-SHORT: MA Dataset Collection

We recruited 5 participants[4], who were each provided a Nokia N95 phone with embedded Python scripts that sampled the accelerometer sensor at 30 Hz for the duration of the study and transferred the collected data to a back-end server. Each user was asked to perform each of the 7 MAs consecutively for $\sim$ 7-10 minutes each, resulting in a per-subject study duration of $\sim$ 50-60 minutes.

As it is impracticably onerous to require each subject to continually record their micro-activity 'ground truth' (e.g., log every change from walk to sit or stand) for extended durations while changing orientation and location of the phone, $User2$-5 collected data with the phone placed only on their *preferred* body position, i.e., where they predominantly carry their phone. We performed an in-depth experimental data collection with $User1$. Specifically, unlike most constrained-usage studies, $User1$ performed each MA while changing the phone's on-body position among: {Shirt Pocket, Pants Front Pocket, Pants Back Pocket}, and while randomly altering the phone's orientation in each of these positions.

We are thus able to study the classification accuracy under the following naturalistic conditions: *a)* when the subject carries the phone in the position (which we know a-priori) that he/she most prefers while performing daily chores, and *b)* when the exact on-body position & orientation of the phone is unknown and subject to random changes.

B. Results of MA Inference

We present the results using a 10-fold cross validation approach (i.e., a 90-10% split of the data). Fig. 2 plots the classification accuracy with various choices of 3D+2D feature vectors for 5 users, with the phone located in each individual user's preferred on-body position (as indicated, this position varies by user). We tested many classifiers[5] on these feature choices. Furthermore, we also applied correlation-based feature selection measure [17] to selectively boost good features. While LibSVM and Adaboost achieved better results, in general, than others, there was no universal winner among the classifiers. Therefore, Fig. 2 plots the average accuracy over all of the classifiers and includes the standard deviation of the accuracy values. The plots correspond to a frame duration of T_f=5 secs.

The plot shows that the classification accuracy is uniformly high across all users (dropping to no lower than 88-89% in the worst-case). We also observe that correlation-based feature reduction, applied on a combined set of $3D_{all}$+$2D_{all}$ features, provides marginally better performance.

Fig. 3 plots the classification accuracy achieved on $User1$ when the orientation of the phone (within each on-body position) changes dynamically. The last column reflects the results when the on-body position of the phone is assumed to be unknown. In this case, the classification is performed and the accuracy is evaluated on the *combined* data of all 3 positions. The figure illustrates the following points: *a)* the MA

[4]Four were university students, one was a co-author

[5]*Decision tree –J48*, *Naive Bayes*, *Bayesian network*, *LibSVM* and *Adaptive Boost* (Adaboost) using J48 as the weak learner

TABLE III: Features Used for Micro-Activity Classification

	Name	*Definition*	*Orientation-independent?*
Feature Components	calibrated 3 axis data (3D)	(x_i, y_i, z_i)	no
	projected 2D (Vertical) [$\vec{p}$]	$\vec{p} = \frac{\vec{d}\cdot\vec{v}}{\vec{v}\cdot\vec{v}} \cdot \vec{v}$, where $v = \langle \bar{x}, \bar{y}, \bar{z} \rangle$ (the mean of x,y,z) and $\vec{d} = \langle x - \bar{x}, y - \bar{y}, z - \bar{z} \rangle$	yes
	projected 2D (Horizontal) [$\vec{h}$]	$\vec{d} - \vec{p}$	yes
	projected 2D (Magnitude) [mag]	$\lvert\vec{h}\rvert, \lvert\vec{p}\rvert, corr(\lvert\vec{h}\rvert, \lvert\vec{p}\rvert)$	yes
Time Domain Features	Mean	AVG($\sum x_i$); AVG($\sum y_i$); AVG($\sum z_i$)	no
	Variance	VAR($\sum x_i$); VAR($\sum y_i$); VAR($\sum z_i$)	no
	Mean-Magnitude	AVG($\sqrt{x_i^2 + y_i^2 + z_i^2}$)	yes
	Magnitude-Mean	$\sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}$	yes
	Two-Axis Correlation	$corr(xy) = \frac{cov(xy)}{\sigma_x, \sigma_y}$; similarly $corr(yz)$, $corr(xz)$	no
	Signal-Magnitude Area (SMA)	$\frac{1}{n}\sum_{i=1}^{n}(\lvert x_i\rvert + \lvert y_i\rvert + \lvert z_i\rvert)$	yes
Frequency Domain Features	FFT Magnitude	$m_j^{(x)} = \lvert a_j + b_j i\rvert$; similarly, $m_j^{(y)}$, $m_j^{(z)}$	no
	FFT Energy	$\frac{\sum_{j=1}^{N}(m_j^2)}{N}$, for x,y,z respectively	no
	FFT Entropy	$-\frac{\sum_{j=1}^{n}(p * log(p))}{n}$, for x,y,z respectively, where p is normalized histogram count of FFT component magnitudes	no

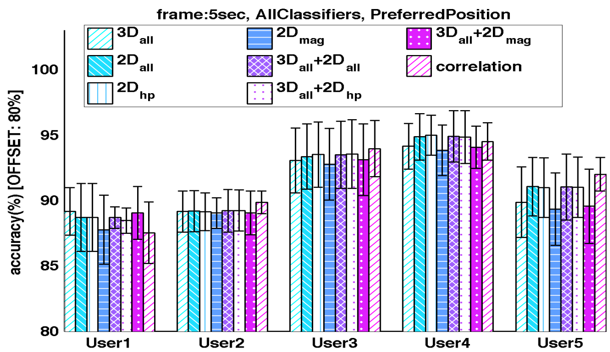

Fig. 2: Micro-Activity Classification accuracy across all 5 users (T_f = 5secs) with the phone in their preferred positions —{front pocket in the pants, back pocket in the pants, shirt pocket in the chest}. The plot shows various combinations of "3D" & "2D" features (ref. Table III). $3D_{all}$ (or $2D_{all}$) implies the use of all time & frequency domain 3D-features (or 2D-features); $2D_{hp}$ refers to features computed on the projected orientation-independent frames – both gravity ($\vec{p}$) and its plane perpendicular ($\vec{h}$) ; $2D_{mag}$ refers to features computed over magnitudes of $\vec{h}$ and $\vec{p}$; 'correlation' refers to the feature selection technique we used.

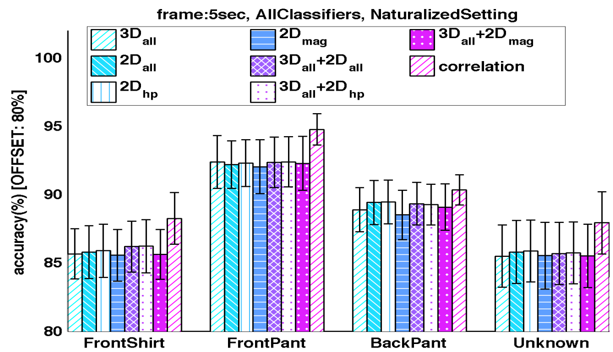

Fig. 3: MA Classification accuracy for *User1* with naturalistic (varying) phone orientations (T_f = 5 secs) and with unknown body positions. 'FrontShirt'=shirt pocket in the chest, 'FrontPants'=front pocket in the pants, 'BackPants'=back pocket in the pants. 'Unknown'=body position is mixed and not given.

classification accuracy is higher when the phone is placed in the lower part of the body (an observation previously made with multiple body-worn sensors [1][8]); *b)* the choice of feature classes result in performance differences of ∼5%, and correlation-based feature selection is the best in classifying such naturalized usage data; *c)* the classification accuracy for the "unknown" case, which best reflects naturalistic usage conditions, is at an acceptably healthy ∼90%; These conclusions hold uniformly across analyses performed with other parameters—e.g., when T_f is varied to be $\{2, 5, 10, 20\}$ secs.

Our experiments establish that SAMMPLE's lower layer can achieve $\sim 90\%$ MA classification accuracy under naturalistic conditions, by applying appropriately-boosted 3D & 2D features, and is thus able to transform the raw data ($\mathcal{A}$) of an unknown HA instance to a reasonably *bona fide* sequence of MA labels. We next describe the feature extraction & classification process used to derive a HA label from this MA sequence. We restrict ourselves to T_f choices of $\{5, 10\}$ secs, as these provided the best MA classification accuracy.

V. Upper Layer: Semantic Activity Inference

We describe two broad approaches for extracting relevant features in SAMMPLE's upper layer, from the sequence of MAs that are associated with each HA instance: *a)* the *Order-Oblivious* (*OO*) and *b)* the *Sequence-Aware* (*SA*) approach. To elucidate each approach, we utilize the illustrative example in Table IV, showing feature extraction by using two types of HA instances (i.e., HA_1 - *Office_Break* and HA_2 - *Office_Lunch*) from training data, with a simple set of MAs (viz. 'walk (w)', 'sit (s)', 'stand (t)').

A. The Order-Oblivious (***OO***) Approach

Given the MA sequence (by MA inference in Section IV) of a HA instance, this approach creates an M-dimensional feature vector (M = No. of MAs), where the i^{th} element of the vector denotes the number of MAs of type MA_i. The feature vector thus captures the duration (as T_f is a constant) of each specific MA in the given HA instance. For example in Column1 of Table IV, the first HA instance's MS sequence $MS_1^{(1)}$ = [t t t t t t w w w w t] has 4 'walk' MAs, 0 'sit' MAs, and 7 'stand' MAs. The corresponding feature-vector for $MS_1^{(1)}$ is $[4, 0, 7]$.

TABLE IV: Running example of feature selection in the Locomotive Signature Space in the 2-$Tier$ approach.

Col. (Column) 1		*Col. 2*	*Col. 3*			*Col. 4*	*Col. 5*	*Col. 6*			*Col. 7*
MA Streams of 2 Types **HA_1: Office_break** **HA_2: Office_lunch** *[w:walk, s:sit, t:stand]*		**OO Features** *[w, s, t]*	**SA-TD Patterns Subseq (size: 3, θ≥0.6)**			**SA-TD Features** *[w, s, t, ttw, tww, wwt, tts, tss, sst]*	**T-P Seq**	**SA-TP Patterns Subseq (size: 3, θ≥0.6)**			**SA-TP Features** *[w, s, t, twt, tst]*
			sub_c	**cov**	**supp**			**sub_c**	**cov**	**supp**	
HA_1	$MS_1^{(1)}$: [tttttwwwwt]	[4, 0, 7]	[ttt] [ttw] [tww] [www] [wwt] [wtt]	1 2 3 1 2 1	1/3 2/3 3/3 1/3 2/3 1/3	[4,0,7,1,1,1,0,0,0]	[twt]	[twt]	2	2/3	[4,0,7,1,0]
	$MS_2^{(1)}$: [tww]	[2, 0, 1]				[2,0,1,0,1,0,0,0,0]	[tw]				[2,0,1,0,0]
	$MS_3^{(1)}$: [ttwwtt]	[2, 0, 4]				[2,0,4,1,1,1,0,0,0]	[twt]				[2,0,4,1,0]
HA_2	$MS_4^{(2)}$: [ttsssst]	[0, 5, 3]	[tts] [tss] [sss] [sst] [wwt] [wtt]	2 2 1 2 1 1	2/2 2/2 1/2 2/2 1/2 1/2	[0,5,3,0,0,0,1,1,1]	[tst]	[tst] [wts]	2 1	2/2 1/2	[0,5,3,0,1]
	$MS_5^{(2)}$: [wwttsst]	[2, 2, 3]				[2,2,3,0,0,1,1,1,1]	[wtst]				[2,2,3,0,1]

*B. The Sequence-Aware (**SA**) Approach*

This approach extracts additional features that capture the *order* (or sequence) in which the various MAs occur within a specific HA instance. This approach should improve discriminatory capability of the resulting features, compared to the OO approach which does not utilize such knowledge. However, it comes at the expense of higher dimensionality of the feature vector. We consider two pattern mining-based techniques to learn such key discriminatory features from the underlying traces: **SA-TD**, a *duration-preserving* strategy and **SA-TP**, a *transition-preserving* strategy. To explain them, we first define a few terms.

Let M_i=$[MS_1^{(i)}, \ldots MS_l^{(i)}]$ be the set of Micro-Activity sequences associated with the l different instances of HA_i. For example, $MS_1^{(1)}$ =[t t t t t t w w w w t] in Table IV. Let $\mathcal{S}_j^{(i)}$ be the set of all *sub-sequences* of $MS_j^{(i)}$. Let sub_c be a micro-activity subsequence that occurs at least once in the combination of all l instances of HA_i, i.e., $sub_c \subseteq \cup_{j=1,\ldots,l} \mathcal{S}_j^{(i)}$.

Definition 1 (Cover of sub_c)**.** *Denoted as* ***cov(**sub_c, M_i**)**, *equals the number of instances in M_i that contains at least one instance of sub_c .*

Definition 2 (Support of sub_c)**.** *Denoted as* ***supp(**sub_c, M_i**)**, *equals* $\frac{cov(sub_c, M_i)}{l}$.

For example, Col.3 in Table IV shows the length-3 subsequence [t t w] occurs 2 times amongst 3 (l=3) instances of HA_1. Hence cov([t t w],HA_1)=2 and supp([t t w],HA_1)=2/3.

Prior work has revealed that patterns with low support in individual classes (each HA label is a class) or with very high support globally across classes are typically not useful for classification, as such patterns either occur very infrequently in the class instances or are a common occurrence in multiple classes, respectively [5]. In our scenario, an individual MA symbol (e.g., 'sit') is likely to be very common in all instances, while a long sequence of MAs will have low *cover*.

There are many types of discriminatory patterns possible in such sequences (e.g., patterns with multiple wild cards) and defining new pattern mining algorithms is not the focus of this paper. Instead we present two broad strategies that augment the OO-based feature vector with additional features defined by *sub-sequences with high observed support.* We choose this because these sub-sequences (sub_cs) should have high discriminatory power w.r.t. at least one other HA in the data.

Temporal Duration-preserving strategy (SA-TD): Given a minimum support threshold Θ_0 and a maximum sub_c size K_{max}, this strategy discovers the set of all sub_cs of length $[2, 3, \ldots, K_{max}]$, that have $supp(sub_c, HA_i) \geq \Theta_0$. For e.g., Col.3 in Table IV shows that the sub_cs of length 3 selected with $\Theta_0 \geq 0.6$ for HA_1 are {t t w}, {t w w}, {w w t}. The SA-TD algorithm finds the *union* of all such qualifying sub-sequences across all HAs in the training data. For e.g., in Table IV, this approach results in the selection of the following 3-element sub_cs as features: {t t w}, {t w w}, {w w t}, {t t s}, {t s s}, {s s t} across all the HAs (HA_1 and HA_2). Intuitively, SA-TD features capture a set of MA transitions among consecutive frames (including self-transitions, i.e., MA sequences of long duration) that are observed to occur often.

The resulting sequence features are appended to the OO features to create a longer OO+$Sequence$ feature vector, with the i^{th} element of the vector corresponding to the frequency of occurrence of the corresponding feature. For example, for the instance $MS_1^{(1)}$ in Table IV, Col.4 shows that the SA-TD approach results in a feature vector $[4, 0, 7, 1, 1, 1, 0, 0, 0]$, where the elements $[1, 1, 1, 0, 0, 0]$ come from the sub-sequence based features, as the sub_cs '[ttw]' & '[wwt]' & '[tww]' occur once each in $MS_1^{(1)}$.

Transition-preserving strategy (SA-TP): This approach preserves only the transitions between *distinct, adjacent* MAs, by removing (or collapsing) the run-length of consecutively repeating MA symbols for each HA instance. E.g., Col.5 in Table IV shows this T-P sequence associated with $MS_1^{(1)}$ is transformed to [t w t]. By focusing purely on the sequence of *transitions* among distinct MAs, the SA-TP approach ignores slight variations in the duration of an individual MA and helps discover key underlying transitions. Consider an activity defined as a 'smoking break'. It is highly likely that two instances of this HA might share a certain latent sequence (e.g., defined by an order of $walk \rightarrow stand \rightarrow walk$), while differing slightly in the duration of each micro-activity (e.g., [w w w t t w] and [w w t t t w w]). The SA-TP approach would not only identify [w t w] as a potential unifying & discriminatory feature, but also help to *reduce the dimensionality* of the resulting feature vector. In our experiments, SA-TP results in $\sim$ 2-4 fold reduction in the size of the resulting feature vector, compared to SA-TD.

Feature Reduction: As the *SA* approaches have lead to high dimensionality of the feature vectors, the feature extraction step is followed by a step of correlation-based selection of good features [17]. Subsequently, HA classification models are built on the final reduced feature space, using labeled training data, to carry out prediction of unknown HA instances.

C. 1-$Tier$ approach

As an alternative to SAMMPLE's 2-$Tier$ approach, the conventional 1-$Tier$ approach extracts time and frequency domain statistical features *directly* from the *raw accelerometer data stream* associated with each HA instance, after the initial step of calibration. The 1-$Tier$ model is common in traditional activity recognition literature [11], [8]. In this approach, for each training instance of HA_i, we consider the accelerometer readings recorded for that instance and compute the statistical orientation-dependent and independent features, just as we did for micro-activity classification (Table III). Thereafter, correlation-based feature selection is applied and a classifier model is built from the training data. HA labels are predicted using the model on unknown test data.

VI. Experimental Results on HA Mining

We now use real-world semantic activity traces to compare the performance of 2-$Tier$ SAMMPLE against the conventional 1-$Tier$ approach.

A. Description of SEMANTIC-LONG: HA Dataset Collection

Our Semantic Activity data collection involved the same 5 users, who volunteered to carry the Nokia N95 smart phone with them[6], in their preferred body position *as they went about performing their daily lifestyle activities*. Alongside, they maintained a separate diary where they tagged *all of the semantic activities* performed, only at their respective office and home locations. This longitudinal data was gathered over a span of 8 weeks on working days, with gaps due to individual variations in lifestyle routines (details in Table V).

TABLE V: Summary of Semantic Activity Dataset

	User 1	*User 2*	*User 3*	*User 4*	*User 5*
No. of Days	27	31	39	32	23
No. of unique HAs	30	64	25	41	65
No. of HA instances	194	215	372	167	228
No. of HA instances selected	186	203	356	165	192

Tagging Process & Principles: The users were provided an initial idea on what constituted a semantic activity 'HA' (e.g., work, break, lunch). Although not mandatory, users often provided additional context for each tag (e.g., break_coffee, break_toilet, office_work_at_desk). Each user recorded the tag tuples: *[activity_start_time, activity_tag]*. As the activities were sequential, the end time of an activity was derived from the start time of the next tag. The last activity performed on a certain day at a certain location had an explicit *end_time* registered by the user. The resulting data (along with a corresponding GPS-based location tag) was periodically transmitted to a back-end server.

In total, we obtained 152 days of data, with each day containing between $\sim$ 4-15 tags/person. Table VI provides some examples from the tag cloud. This data was cleaned by applying a per-user manual process of normalization and information summarization: (1) Semantically equivalent tags (e.g., office_meet and office_meeting) were converted to a standard notation. (2) Tags having additional context were collapsed to the corresponding root tag (e.g., office_meet_colleague $\rightarrow$ office_meet), unless the activity occurred sufficiently frequently, and vice versa–e.g., office_break_toilet was separated from office_break for some users. Infrequent tags were subsequently removed from further investigation (e.g., home_freshenup). This resulted in a total of 1102 HA instances across all users.

[6]Four users used it as their primary cellphone for the duration of our data collection; only one user is in the author list of this paper.

TABLE VI: Examples showing Tag Clouds collected (right column) and corresponding normalized Tags (left column)

HA Label	Examples of User Tags
O_work	office_work, work_work, office_work_TA, office_work_check_printer
O_break	office_break, office_break_walk around office_break_talk,
O_coffee	office_coffee_break, office_break_tea, office_short_break_coffee
O_toilet	office_break_toilet, work_break_toilet, office_short_break_toilet
O_meet	office_meet, office_meet_lab, office_meeting, office_meet_NRC
O_lunch	office_lunch, work_lunch, office_lunch_desk, office_break_lunch
H_work	home_work, home_work_move, home_work_on_computer
H_relax	home_relax, home_relax_freshen_up, home_relax_movie
H_break	home_break, home_break_shopping, home_break_coffee
H_cook	home_cook, home_cooking, home_clean_dishes, home_wash_dishes
H_eat	home_eat, home_lunch, home_dinner, home_eat_adults_with_movie
H_baby	home_baby_routine, home_baby_routine_eat_with_baby

Data Processing: The continuous accelerometer stream $\mathcal{A}$ was first calibrated appropriately (described earlier for MA classification). Subsequently, the (start, end) timestamps of each HA instance were used to extract the corresponding accelerometer segment. The individualized MA classification model, derived in Section IV, was then applied to generate the $\mathcal{MS}$ stream for each HA instance. The segmented $\mathcal{A}$ and $\mathcal{MS}$ streams were then used to learn the HA classification models for the 1-$Tier$ and 2-Tier approaches respectively. As the number of HA instances is lower than the corresponding number of MA instances (one HA consists of a sequence of MAs), the HA classification accuracy was computed using an 8-fold cross-validation approach (i.e., a 87.5-12.5% split of the data), rather than the 10-fold validation employed earlier for classifying MAs.

B. Performance Comparison of Approaches

Fig. 4 presents the classification accuracy statistics observed in 1-$Tier$ method and the four feature extraction strategies in the 2-$Tier$ SAMMPLE approach (i.e., OO, SA-TD, SA-TP and the combined feature sets SA-TD+TP). We experimented with a variety of classifiers (*decision tree –J48*, *Adaptive Boost - Adaboost*, *LibSVM*, *Bayesian Network* and *Naive Bayes*) and plot the mean, and the standard deviation of the measured accuracies across these classifiers. Due to the lack of space, we only show three users in Fig. 4; but later provide the details of the comparative confusion matrices for all users (in Section VI-C).

We observe that SAMMPLE results in an across-the-board improvement in the classification accuracy, ranging from 7-20%, compared to the 1-$Tier$ approach. Due to the different dynamics of lifestyle activities of different users, the absolute accuracy values are user-dependent. A salient observation is that even the OO approach, with a slim feature vector dimension (7 MAs), mostly out-performs the 1-$Tier$ approach which uses $\sim$ 70 statistical features (with correlation-based feature selection). We also note that sequence-based features provide

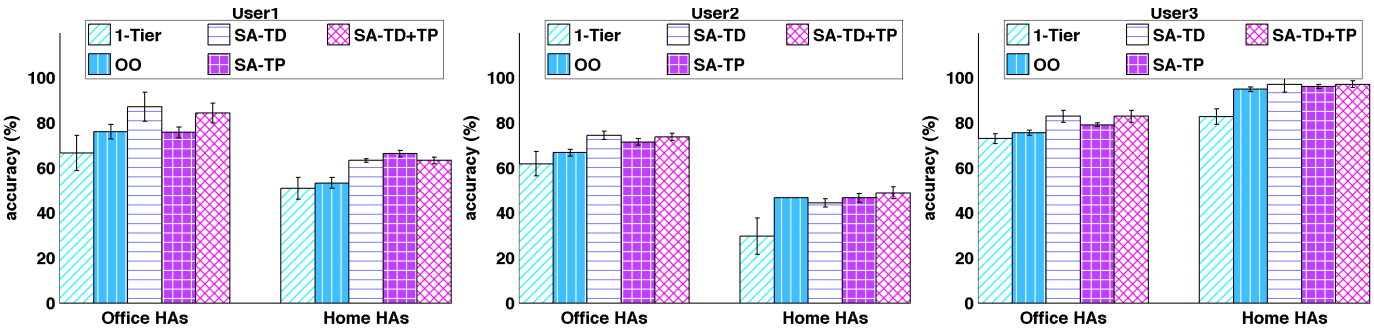

Fig. 4: Performance comparison between 1-$Tier$ approach and SAMMPLE ($K_{max} = 4$; $\Theta_0 = 0.7$; MA frame-size $T_f = 5secs$). The bars indicate the mean accuracy values obtained across all classifiers and also show the standard deviation of the accuracies. 'Office HAs' and 'Home HAs' indicate office and home semantic activities.

an additional, but variable (4-15%), amount of improvement in the classification accuracy. These results establish the superior quality of locomotive signatures, compared to their statistical counterparts. Intuitively, the 1-$Tier$ statistical feature-based approach should provide almost comparable performance to SAMMPLE, for those semantic activities that are dominated by a single locomotion. We observe this in some HAs that have dominating "sitting" states, like *User*1's office_work.

Between SA-TD and SA-TP, there seems to be no clear winner across all users. In Fig. 5, we perform a sensitivity analysis of the algorithms[7] with K_{max} (the maximum possible sequence length considered in SA-TD & SA-TP) varying from 2 to 6. We observe that the classification accuracy shows a marginal improvement initially, before flattening out at K_{max} beyond 3 or 4. This demonstrates that relatively-short MA sequences possess the highest discriminatory power. Note that sequence-based features have greater discriminatory power when different instances of the same HA share a common sequence of HA transitions. Accordingly, the improvement in HA classification accuracy via the SA-TD and SA-TP approaches is also activity and user-dependent, as different users might perform the same daily activity (e.g., office_break) in predictable or random fashion.

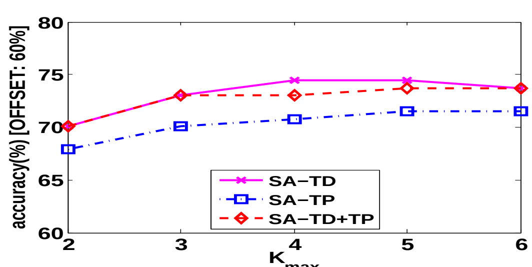

Fig. 5: Sensitivity Analysis of K_{max}

C. Confusion Matrices for Activity-level Results

Table VII provides the comparative confusion matrices of the $1-Tier$ and 2-$Tier$ approaches. Overall, we observe that the 2-$Tier$ approach increases the HA classification accuracy to a healthy $\geq$ 75-90% for several semantic activities. We can clearly observe differences in accuracy gains between home and office activities. As shown in the last row of Table VII, across all of the 1102 instances, the accuracy of 2-$Tier$ is 77.14%, with the gain of 16.37%, compared to 60.77% by 1-$Tier$. The performance improvement is prominent for activities like O_break, with some users (e.g., user 4 and 5) recording $\geq$ 20% per-activity accuracy gain.[8] This corroborates our observation that SAMMPLE's approach is better, than the statistical feature-based approach, in classifying those activities that have a richer mix of locomotive states. Overall, the results demonstrate that SAMMPLE can provide improved accuracy for smartphone-based semantic activity classification in real-world scenarios.

Compared to *Office* activities, *Home* activities generally seem to have higher degrees of confusion, independent of the performance gains achieved by the 2-Tier approach. Some interesting anomalies are seen (e.g., H_eat for *User4*, H_eat for *User5*) when performance drops in the 2-Tier approach. This could be due to several reasons – domination of a single locomotive state, or un-balanced activity instances for learning the models. These results suggest that additional sensory context (e.g., microphones) might be needed to obtain enough discriminatory features to disambiguate among such HAs.

VII. Conclusions and Future Work

We proposed & evaluated SAMMPLE, a 2-$Tier$ process for extracting and applying locomotion-based discriminatory features to classify semantic activities in un-instrumented environments, using data from a single phone-embedded accelerometer. Using longitudinal lifestyle data (5 users for 152 days), we compared this approach with the traditional statistical feature-driven classification (1-$Tier$) approach, and showed that SAMMPLE provided an overall accuracy improvement of 16.37% (from 60.77% to 77.14%), based on the total set of 1102 semantic activity instances.

While this paper investigated the limits of the discriminatory power of accelerometers, SAMMPLE is a generic methodology that can be extended to incorporate additional sensors (e.g., microphones), which will undoubtedly improve the classification accuracy. In the future, we plan to incorporate such additional sensors, and expand our scope to include unsupervised learning of the activity vocabularies under such naturalized settings.

[7] Results shown only for $User2$ due to lack of space. Similar results were observed for the other 4 users.

[8] *User3* is an unusual case, with his O_break not being detected well by both approaches, due to the high degree of observed similarity in executing the different breaks. Along similar lines, we note that O_work and O_meet are also sometimes confusing.

TABLE VII: Confusion Matrices for all 5 users providing activity-wise splits of accuracies. Comparison is given between 1-$Tier$ and SAMMPLE's 2-$Tier$ approach. An entry in the i^{th} row & j^{th} column denotes the fraction of HA_i instances where the classifier predicted the activity as HA_j. The bounding boxes indicate the HAs recording accuracy increase $\geq$10% by using SAMMPLE. Accuracy drops are underlined. Last column indicates average accuracy gain per activity cluster in home and office.

	HA	No.	1-Tier Method Confusion Matrix	& Avg. Accuracy	2-Tier Method (SAMMPLE) Confusion Matrix	& Avg. Accuracy	Gain
USER 1	O_work	30	.900 .100 .000 .000	57.53%	.900 .033 .067 .000	87.14%	29.61%
	O_break	17	.176 .824 .000 .000		.000 .941 .059 .000		
	O_meet	15	.467 .467 .067 .000		.143 .071 .786 .000		
	O_lunch	11	.000 .818 .182 .000		.000 .111 .111 .778		
	H_work	36	.833 .033 .100 .000 .033	45.54%	.931 .000 .000 .034 .034	66.33%	20.79%
	H_cook	21	.143 .619 .190 .048 .000		.000 .789 .158 .000 .053		
	H_relax	25	.053 .526 .316 .000 .105		.000 .316 .368 .105 .211		
	H_break	14	.071 .571 .214 .000 .143		.000 .071 .143 .429 .357		
	H_eat	17	.176 .000 .471 .000 .353		.118 .059 .000 .235 .588		
USER 2	O_work	80	.788 .075 .100 .038	59.03%	.947 .018 .018 .018	74.79%	24.76%
	O_break	33	.167 .667 .125 .042		.125 .792 .042 .042		
	O_meet	20	.750 .200 .050 .000		.526 .158 .158 .158		
	O_lunch	19	.550 .100 .100 .250		.211 .053 .053 .684		
	H_work	9	.333 .000 .667 .000	30.43%	.000 .889 .111 .000	48.89%	18.46%
	H_eat	12	.333 .000 .667 .000		.091 .727 .182 .000		
	H_relax	21	.313 .000 .688 .000		.000 .125 .875 .000		
	H_baby	9	.333 .000 .667 .000		.000 .333 .667 .000		
USER 3	O_work	122	.902 .025 .041 .008 .025	71.73%	.941 .000 .034 .008 .017	82.91%	11.18%
	O_break	15	.133 .067 .400 .267 .133		.133 .000 .267 .333 .267		
	O_coffee	35	.171 .057 .486 .257 .029		.171 .000 .686 .143 .000		
	O_toilet	33	.121 .152 .182 .545 .000		.030 .000 .091 .848 .030		
	O_lunch	32	.125 .031 .031 .063 .750		.031 .000 .031 .000 .938		
	H_eat	28	.929 .000 .000 .071	83.19%	1.000 .000 .000 .000	97.14%	13.95%
	H_relax	57	.000 .965 .000 .035		.018 .982 .000 .000		
	H_clean	11	.000 .000 .636 .364		.000 .000 1.000 .000		
	H_cook	23	.087 .130 .304 .478		.111 .000 .111 .778		
USER 4	O_work	59	.932 .000 .068	69.57%	.964 .018 .018	86.79%	17.22%
	O_meet	15	.600 .000 .400		.267 .600 .133		
	O_break	41	.390 .000 .610		.111 .056 .833		
	H_work	7	.286 .571 .143 .000	52.00%	.857 .000 .143 .000	67.39%	15.39%
	H_eat	18	.111 .778 .056 .056		.167 .556 .278 .000		
	H_relax	15	.200 .467 .333 .000		.462 .000 .538 .000		
	H_cook	10	.100 .100 .300 .500		.000 .000 .000 1.000		
USER 5	O_work	65	.723 .031 .092 .138	52.78%	.937 .000 .063 .000	70.99%	18.21%
	O_meet	11	.273 .000 .636 .091		.545 .000 .364 .091		
	O_break	45	.222 .089 .600 .089		.132 .000 .868 .000		
	O_lunch	23	.565 .000 .348 .087		.526 .000 .421 .053		
	H_work	6	.000 .667 .000 .333	41.60%	.000 .167 .167 .667	52.27%	10.60%
	H_eat	11	.000 .273 .000 .727		.000 .000 .000 1.000		
	H_relax	11	.000 .091 .000 .909		.000 .000 .600 .400		
	H_cook	20	.000 .150 .000 .850		.000 .000 .000 1.000		
Total HA No.		**1102**	Avg. Accuracy (all HAs):	**60.77%**	Avg. Accuracy (all HAs):	**77.14%**	**16.37%**

REFERENCES

[1] L. Bao and S. Intille. Activity Recognition from User-Annotated Acceleration Data. In *Pervasive*, pages 1–17, 2004.

[2] T. Brezmes, J.-L. Gorricho, and J. Cotrina. Activity Recognition from Accelerometer Data on a Mobile Phone. In *IWANN (2)*, pages 796–799, 2009.

[3] A. Campbell, S. Eisenman, N. Lane, E. Miluzzo, R. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G.-S. Ahn. The Rise of People-Centric Sensing. *IEEE Internet Computing*, 12:12–21, 2008.

[4] X. Cao, G. Cong, and C. S. Jensen. Mining Significant Semantic Locations From GPS Data. In *VLDB*, pages 1009–1020, 2010.

[5] H. Cheng, X. Yan, J. Han, and C. Hsu. Discriminative Frequent Pattern Analysis for Effective Classification. In *ICDE*, pages 716–725, 2007.

[6] T. Choudhury and et. al. *IEEE Pervasive Computing*, 7(2):32–41, 2008.

[7] D. Choujaa and N. Dulay. Predicting Human Behaviour from Selected Mobile Phone Data Points. In *Ubicomp*, pages 105–108, 2010.

[8] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu. epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition. In *PerCom*, pages 1–9, 2009.

[9] J. Healey and B. Logan. Wearable wellness monitoring using ecg and accelerometer data. In *ISWC*, pages 220–221, 2005.

[10] T. Hunh, M. Fritz, and B. Schiele. Discovery of Activity Patterns using Topic Models. In *UbiComp*, 2008.

[11] T. Kasteren, A. Noulas, G. Englebienne, and B. Krose. Accurate Activity Recognition in a Home Setting. In *Ubicomp*, pages 1–9, 2008.

[12] N. Kern and B. Schiele. Context-aware notification for wearable computing. In *ISWC*, pages 223–230, 2003.

[13] H. Lu, W. Pan, T. Choudhury, and A. T. Campbell. Soundsense: Scalable sound sensing for people-centric applications on mobile phones. In *MobiSys*, pages 165–178, 2009.

[14] H. Lu, Yang, J. Liu, N. Lane, T. Choudhury, and A. Campbell. The Jigsaw Continuous Sensing Engine for Mobile Phone Applications. In *Sensys*, pages 71–84, 2010.

[15] E. Miluzzo, N. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. Eisenman, X. Zheng, and A. Campbell. Sensing Meets Mobile Social Networks: the Design, Implementation and Evaluation of the Cenceme Application. In *Sensys*, pages 337–350, 2008.

[16] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity Recognition from Accelerometer Data. In *AAAI*, pages 1541–1546, 2005.

[17] L. Smith and M. Hall. Feature Subset Selection: a Correlation Based Filter Approach. In *ICONIP*, 1997.

[18] E. Tapia, S. Intille, and K. Larson. Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors. In *Pervasive*, pages 158–175, 2004.

[19] T. T. Truyen, H. H. Bui, D. Q. Phung, and S. Venkatesh. Learning Discriminative Sequence Models from Partially Labelled Data for Activity Recognition. In *PRICAI*, pages 903–912, 2009.

[20] Y. Wang, Q. Jacobson, J. Lin, J. Hong, N. Sadeh, M. Annavaram, and B. Krishnamachari. A framework of energy efficient mobile sensing for automatic user state recognition. In *MobiSys*, pages 179–192, 2009.

[21] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer. SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories. In *EDBT*, pages 259–270, 2011.

[22] J. Yang, H. Lu, Z. Liu, and P. Boda. Physical Activity Recognition with Mobile Phones: Challenges, Methods, and Applications. In *Multimedia Interaction and Intelligent User Interfaces. Advances in Pattern Recognition*, pages 185–213, 2010.

[23] T. yu Wu, J. Y. jen Hsu, and Y. ting Chiang. Continuous recognition of daily activities from multiple heterogeneous sensors. In *AAAI*, pages 80–85, 2009.

[24] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative Location and Activity Recommendations With GPS History Data. In *WWW*, pages 1029–1038, 2010.

[25] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining Interesting Locations and Travel Sequences from GPS Trajectories. In *WWW*, pages 791–800, 2009.