An Exploration with Online Complex Activity Recognition using Cellphone Accelerometer

Zhixian Yan[a]
Samsung Research, USA
zhixian.yan@samsung.com

Dipanjan Chakraborty
IBM Research, India
cdipanjan@in.ibm.com

Sumit Mittal
IBM Research, India
sumittal@in.ibm.com

Archan Misra
Singapore Management University
archanm@smu.edu.sg

Karl Aberer
EPFL, Switzerland
karl.aberer@epfl.ch

[a]This work was performed while the author was at EPFL.

UbiComp'13 Adjunct, September 8–12, 2013, Zurich, Switzerland.
ACM 978-1-4503-2215-7/13/09.
http://dx.doi.org/10.1145/2494091.2494156

Abstract
We investigate the problem of *online* detection of complex activities (such as *cooking*, *lunch*, *work at desk*), i.e., recognizing them while the activities are being performed using parts of the sensor data. In contrast to prior work, where complex activity recognition is performed *offline* with the observation of the activity available for its entire duration and utilizing deeply-instrumented environments, we focus on online activity detection using only accelerometer data from a single body-worn smartphone device. We present window based algorithms for online detection that effectively perform different tradeoffs between *classification accuracy* and *detection latency*. We present results of our exploration using a longitudinally-extensive and clearly-annotated cellphone accelerometer data trace that captures the true-life complex activity behavior of five subjects.

Author Keywords
activity recognition, complex activities, accelerometer, online detection, detection accuracy, detection delay

ACM Classification Keywords
I.2.1 [Applications and Expert Systems]

General Terms
Algorithms, Experimentation

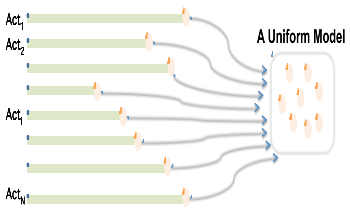

Figure 1: Uniform window model

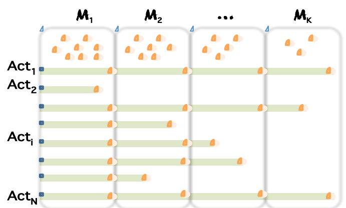

Figure 2: Multiple window model

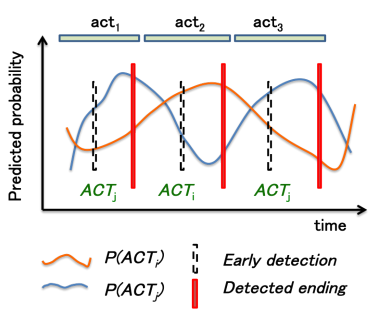

Figure 3: Continuous ACT detection – three activity boundaries detected based on this strategy. The three activities (act_1,act_2,act_3) are detected as ACT_j, ACT_i, and ACT_j, respectively.

Introduction

Complex activities are referred to as *Activities of Daily Living* (ADL) that can be detected using sensor data collected from a consumer-grade personal mobile device like smartphone. Examples of complex activity are "cooking at home", "watching TV at home", "lunch in office". Prior work on such complex activity detection operates *offline* [2, 4], which means the classification is performed by analyzing the entire stream of sensor data after the completion of an activity episode, based on features extracted over the entire activity duration. Additionally, it utilizes *deeply-instrumented settings* [1, 3] such as the use of multiple body-worn accelerometer sensors or RFID tagging of household objects. Our work differs from such works from three critical aspects: (1) *online or in-situ complex activity detection*, i.e., the ability to classify ADLs rapidly and early, *while they are being performed*; (2) require no special instrumentation of either the individual or the environment, and instead investigate the extent to which such online classification can be performed solely using accelerometer data from a *commodity smartphone*; (3) perform the algorithms under natural conditions of daily living, i.e., *in the wild*.

Our motivation is to investigate the research challenges of online continuous complex activity detection in completely naturalized real-life data, with a single smartphone accelerometer. The detailed contributions are:

1. We propose two different window-based learning approaches for *early detection* of complex activities. By investigating discriminatory features of different complex activities, we study their across-time effects both on *classification accuracy* and *detection delay* for online detection.
2. To perform continuous activity recognition in a streaming fashion, we present a simple, but demonstrably effective, *demarcation algorithm* that identifies the points where an individual transition from one complex activity to another.
3. We evaluate our approach using a real-life naturalized setting dataset that is significant both for its longitudinal trace and for the availability of *ground-truth* of corresponding complex activities.

Online Continuous Approach

Our approach is designing window-based model to study early detection of complex activities in real-time using cellphone accelerometer; in addition, we investigate the continuity of accelerometer streams for detecting activity transitions (e.g., from 'cook' to 'eat').

Window based Early Detection Strategies

We design window based strategies to make *early, online* detection of activities from accelerometer streams. Before proceeding to describe our algorithms, it is first important to establish the set of *classificatory features*, defined over each frame of observed accelerometer data. A frame is a small window of accelerometer data, e.g., 5 secs or 10 secs. Like [4], we transform the raw accelerometer stream into locomotive features - each frame indicates a motion status, such as *sit*, *stand* and *walk*, in order to get a rich set of feature inputs. We design two window models.

- *Uniform Modeling* – The first model is building a *single* classifier model $\mathcal{M}$, which is trained by using the *entire duration* of the complex activity instances (see Fig. 1). For each instance of a specific complex activity Act_i, we calculate a set of features computed over the *entire duration of the activity instance*. All activities are used to train a uniform model that will be used for online testing with partial ongoing sensor streams.

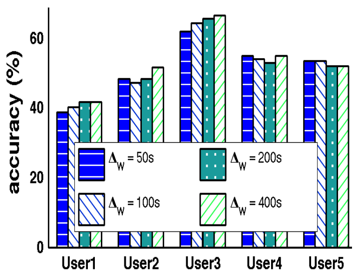

Figure 4: Detection accuracy uniform model (δ_{conf}=0.9)

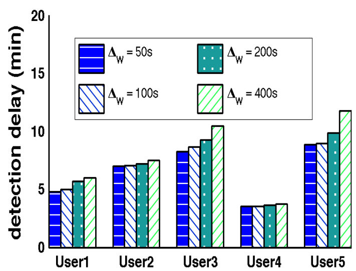

Figure 5: Detection delay with uniform model (δ_{conf}=0.9)

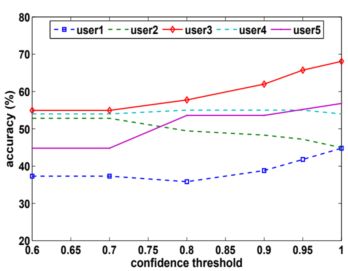

Figure 6: Accuracy vs. Confidence threshold (uniform model, Δ_W = 50S)

- *Multiple-Window based Models* – Complex activities typically have pronounced differentiation during certain sub-intervals. Intuitively, classification could be accurate if we had explicit and separate models that were trained on the corresponding *time interval* of the activities. Accordingly, we investigate the use of 'multiple-window' based approach, where each time window is associated with a different classifier model that is tuned to the specific features exhibited by the complex activities within that window. We build K separate models (i.e., $\mathcal{M}_1$ to $\mathcal{M}_K$ in Fig. 2), corresponding to the time-window partitions of an ACT. The online testing is incrementally using the K models that denote the time-instants that partition the incoming online sensor data stream.

Online Demarcation Strategies
After early detection of activity label, we need to identify the activity duration, i.e., (ts_i, te_i) from a continuous stream containing a sequence of complex activities. Fig. 3 provides a sketch of our online activity boundary segmentation algorithm. For simplicity, we show the real-time prediction probabilities of two activity labels, i.e., $p(ACT_i)$ and $p(ACT_j)$. Since the activities are sequential, the *end* time of the current activity also indicates the *start* time of the next one.[1] Consider a new activity has started at t_{start} and is predicted as ACT_i at t_{detect}. As the stream continues, we calculate the complete prediction vector of all possible activities $predVec = \langle \cdots, p(ACT_i), \cdots, p(ACT_j), \cdots \rangle$ at each time t_{curr}. We can declare that t_{curr} is the t_{end} of ACT_i if the following conditions hold:

[1]Although concurrent activities can be also supported using the probabilistic model, this is not the focus of this paper.

- *Persistent Gradient Monotonicity* - δ_{trend}: The probability of the currently predicted activity, i.e., $p(ACT_i)$, shows negative slope, and in the meanwhile, *at least* another activity probability [i.e., $p(ACT_j)$, $j \neq i$] shows positive slope, for a time duration larger than δ_{trend}.
- *Exceed Minimal Duration* - $timeOut(ACT_i)$: For each activity type, a timeout value is set to the minimum duration of all instances of that ACT. This means that we cannot declare 'ACT_i has ended' until at least the corresponding timeout (ACT_i) has elapsed. Hence, a transition can be declared only if the current time instant t_{curr} satisfies the minimal duration property, i.e., t_{curr}-$t_{start} \geq timeOut(ACT_i)$.

Activity Label Refinement
So far, we have computed activity labels and activity boundaries only from sensor observations at current status. Once the end time of the current activity ACT_i has been predicted, we exploit an activity *state transition matrix* between different activities, i.e., $p(ACT_i|ACT_j)$, to re-compute the predicted label for ACT_i. In effect, this is a *refinement* of the predicted label from the early detection strategies, with knowledge of the activity evolution till its (predicted) boundary and knowledge of the transition probabilities. The transition matrix $p(ACT_i|ACT_j)$ is computed from the available ground truth data. We recompute a new $p(\widehat{ACT_i^{(t)}})$ after the boundary of ACT_i being predicted, combining the support of the current observations in time state t and the state transition from the previous activity at time state *t-1*.

$$p(\widehat{ACT_i^{(t)}}) = w_1 * p(ACT_i^{(t)}) + w_2 * p(ACT_i^{(t)}|ACT^{(t-1)})$$

$$= w_1 * p(ACT_i^{(t)}) + w_2 * \sum_{j=1}^{N} \{p(ACT_j^{(t-1)}) \times p(ACT_i|ACT_j)\}$$

		I. - Ground Truth		II. No Refinement $(p(ACT_i^{(t)}))$			III. With Refinement $(p(\widehat{ACT}_i^{(t)}))$		
		#stream	#realACT	#predACT	TimeSlice (100%)	Offset (min)	#predACT	TimeSlice (100%)	Offset (min)
Home	User 1	12	84	35	0.40145	35	26	0.35459	26
	User 2	8	31	21	0.32134	1.75	21	**0.37361**	10.5
	User 3	11	73	48	0.51284	9.6	23	**0.54236**	23
	User 4	19	42	40	0.34631	5	39	**0.38006**	19.5
	User 5	10	22	21	0.36077	3	24	**0.53613**	4
Office	User 1	7	63	30	0.36873	30	37	0.28781	37
	User 2	8	75	103	0.48676	8.58	107	0.44864	8.2308
	User 3	19	105	94	0.54413	18.80	92	**0.55400**	18.4
	User 4	15	93	47	0.62077	5.88	52	**0.67577**	7.4286
	User 5	14	97	67	0.43107	9.57	72	**0.48039**	6.5455

Figure 7: Online demarcation performance

Exploration Results

Our experiments are based on previous real-life 'in-the-wild' accelerometer data collected from 5 users carrying a Nokia N95 phone, loaded with an application that sampled the accelerometer at 30Hz, continuously as the users were immersed in their daily activities at home or in office. The data details can be found in [4].

Window-based Early Detection – We experimented both uniform and multiple-window based models, for testing how models built with partial observation data. We identified the accuracy increases with increasing frame increments (Δ_W), while an overly large value of Δ_W can cause more errors in transition detection, especially if the incremented buffer straddles the boundary of two activities and includes a bigger chunk of the next activity. Fig. 4-6 shows the early detection performance for all 5 subjects. We found that most complex activities can be detected within ≈5-25% of their overall activity length. In essence, many of the long running activities (30-100 mins) could be detected within 4-20 minutes of their onset. A notable aspect was that almost all activities need less than 50% of their data, in order to be classified with the same accuracy as they would be with complete data.

Demarcation detection and label refinement – Fig. 7 shows the detailed continuous detection performance of five users. We fixed the monotonicity tolerance δ_{trend} to 10 for the state transition algorithm and $timeOut(ACT_i)$ was computed from activity logs. We consider two metrics, $Timeslice$ and $Offset$. The $Timeslice$ accuracy is the percentage of the stream that has been correctly predicted by the algorithm; and the $Offset$ is the time difference between the predicted activity transition and real transition. Fig. 7 shows better performance of using label refinement compared to the inference only using current status sensor observations. The results are indicative of the overall performance that the accelerometer alone as a sensor can achieve in real settings. Our results give cues on which activities are confusing in real-life settings. One can investigate alternate sensors on the phones (e.g. microphone) for such confusing activities.

Conclusion

This paper explored a challenging problem of continuous online detection of complex activities *in the wild* using only a single accelerometer sensor embedded in the smartphone. We designed both uniform and multiple window-based models to study early detection of complex activities in real-time, and demarcation method to detect activity transitions. We showed exploration results of *detection accuracy* and *detection delay*, and also analyzed the tradeoffs between the two metrics.

References

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

[2] Huynh, T., Fritz, M., and Schiele, B. Discovery of activity patterns using topic models. In *Ubicomp* (2008), 10–19.

[3] Wu, T., Chiang, Y., and Hsu, Y.-j. Continuous recognition of daily activities from multiple heterogeneous sensors. In *AAAI Human Behavior Modeling* (2009), 80–85.

[4] Yan, Z., Chakraborty, D., Misra, A., Jeung, H., and Aberer, K. Sammple: Detecting semantic indoor activities in practical settings using locomotive signatures. In *ISWC* (2012), 37–40.