On Strategies for Budget-Based Online Annotation in Human Activity Recognition

Tudor Miu
School of Computing Science
Newcastle University
Newcastle-upon-Tyne, UK
NE1 7RU
* t.a.miu@ncl.ac.uk

Paolo Missier
School of Computing Science
Newcastle University
Newcastle-upon-Tyne, UK
NE1 7RU
paolo.missier@ncl.ac.uk

Daniel Roggen
Sensor Technology Research Centre
School of Engineerng & Informatics
University of Sussex, UK
BN1 9QT
D.Roggen@sussex.ac.uk

Thomas Plötz
Culture Lab
School of Computing Science
Newcastle University
Newcastle-upon-Tyne, UK
NE1 7RU
thomas.ploetz@ncl.ac.uk

UbiComp '14, September 13 - 17 2014, Seattle, WA, USA
Copyright 2014 ACM 978-1-4503-3047-3/14/09. . . $15.00.
http://dx.doi.org/10.1145/2638728.2641300

Abstract
Bootstrapping activity recognition systems in ubiquitous and mobile computing scenarios often comes with the challenge of obtaining reliable ground truth annotations. A promising approach to overcome these difficulties involves obtaining online activity annotations directly from users. However, such direct engagement has its limitations as users typically show only limited tolerance for unwanted interruptions such as prompts for annotations. In this paper we explore the effectiveness of approaches to online, user-based annotation of activity data. Our central assumption is the existence of a fixed, limited budget of annotations a user is willing to provide. We evaluate different strategies on how to spend such a budget most effectively. Using the Opportunity benchmark we simulate online annotation scenarios for a variety of budget configurations and we show that effective online annotation can still be achieved using reduced annotation effort.

Author Keywords
budget-based annotation, activity recognition, online learning

ACM Classification Keywords
I.6.m [Simulation and Modelling]: Miscellaneous.

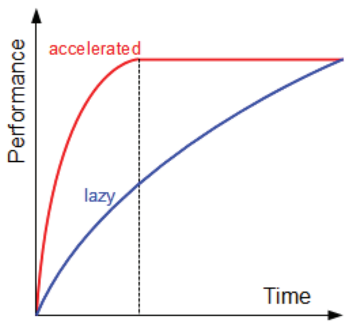

Figure 1: Influence of annotation strategies on online activity recognition systems (schematic): Accelerated (red) vs. slower learning (blue).

Introduction

As an enabling technology, automatic inference of the activities humans are engaged in plays a central role in the majority of ubiquitous and mobile computing applications. Targeting real-world scenarios, Human Activity Recognition (HAR) techniques are often developed in “field deployments”, i.e., keeping prospective users in the loop from early stages of the development process. For example, model personalization is of importance for healthcare settings, which requires individual input to generate personalized feedback during physical exertion [5], or to target individual medical conditions [4, 14]. User involvement becomes a technical necessity, where models need to be adapted or even bootstrapped ’from scratch’, without having access to prior knowledge.

In contrast to lab-based developments, in such contexts it is often difficult to obtain ground truth annotations required for deriving automatic recognizers. Reasons for this can be of a very practical nature as it is often simply impossible to follow a user of mobile HAR technology for the sake of labelling sample data. More importantly, ethical restrictions often prevent direct observations aimed at obtaining ground truth annotations such as in private (smart) homes. Alternative annotation strategies engage users directly, e.g., through self-reporting of activities [9], or through experience sampling, i.e., prompting users to provide labels for current or previous activities [2]. Such user involvement is disruptive as it interferes with ongoing activities with what appears to be mundane support for a technical system – a task that is typically not the primary focus of the user. Arguably, the tolerance for such active user participation is thus limited.

In this paper we focus on optimising prompting in online annotation scenarios. Especially for bootstrapping HAR systems this is a non-trivial endeavour since existing approaches, such as active learning [8] or semi-supervised learning [20], are not applicable as they require prior knowledge about activities to be recognised, i.e., existing annotated data for estimating the underlying distributions.

Our central hypothesis is the existence of a *fixed budget of user provided annotations* that a HAR system may spend during its bootstrapping phase. Spending one unit of the budget corresponds to asking the user for the label of an activity. Focusing on online annotation we assume a “worst memory” scenario where users can only provide information regarding their most recent activity. A request for annotation can be made at any given time as long as there is budget available. We also assume users labels are always reliable and correct. With these assumptions we explore the effectiveness of possible budget allocation strategies. We explore how annotation strategies impact *model performance*. Fig. 1 illustrates the ideal trade-off, where how a greedy strategy (red) translates in more accelerated model performance than a uniform strategy (blue). While accelerated learning (red) results in a reliable model earlier on, this may come at a cost in terms of aggravated *user tolerance* which may impact further interaction with the device. The lazier strategy (blue) learns more slowly but might be preferred by the user in the long run.

The main contribution of this paper is an experimental exploration of various configurations and trade offs between budget levels and spending strategies on one side, and accuracy of the HAR models that can be learned in such settings, on the other. Our findings serve as guidelines for designers of interactive online annotation interfaces to support them in user-centred studies. Specifically, we develop and evaluate budget-based strategies for online annotation of HAR by means of an extensive case study where we simulate online annotation scenarios. We use the Opportunity challenge dataset [7], which comprises of non-periodic activities recorded using multiple worn sensors. We use this realistic simulation to study different problem configurations in detail and in an objective and reproducible way.

Our findings suggest that effective online annotation of human activities can be best achieved using a deterministic greedy budget spending strategy or a probabilistic strategy employing an exponential distribution function. Furthermore, the proposed approach allows us to suggest realistic budget sizes for online annotation tasks. Given that Opportunity is regarded as a realistic and at the same time challenging HAR dataset, these findings are very encouraging for related real-world deployments of budget-based online annotation, including one of ours which is currently in development.

Background

Although supervised learning is the predominant approach for HAR applications, obtaining sufficiently reliable ground truth annotations for training data is challenging, largely due to practical as well as ethical reasons, especially in mobile and domestic ubiquitous computing settings. One approach to overcome these difficulties involves periodically engaging users in the data collection process, by prompting them to label their activities as they occur [10]. Though the general approach is straightforward, one needs to balance the need to acquire new annotations, with limited users' tolerance to interruption of what they perceive as unmonitored activities.

User Perspective

Tolerance to interruption while performing an activity is naturally idiosyncratic, as personal patience and inclination to collaborate come into play. Clearly explaining to the users the purpose of the interruptions and the benefits that can be expected is one way to try and increase their tolerance. Key acceptance factors include the nature of the task, and user awareness that a device is gathering information about the task itself. In this context, the notion of *intelligibility* has been adopted by the ubiquitous computing community to measure and improve upon the capabilities of interactive systems [11].

Technical Perspective

Whilst our research is informed by the user perspective as described above, our main concern is with the technical challenges associated with online user annotation. A number of strategies have been developed for engaging users in the process of activity labelling. For example, the *Experience Sampling Method (ESM)*, also known as *Ecological Momentary Assessment (EMA)* [18], is a technique to prompt users to repetitively reflect upon a relevant current state of fact while, at the same time, functioning normally in their environment. In activity recognition, ESM/EMA has been used to monitor users in naturalistic environments, and to interrupt them in order to annotate some of their activities [10]. Gathering annotations directly from the user eliminates the need for external supervision.

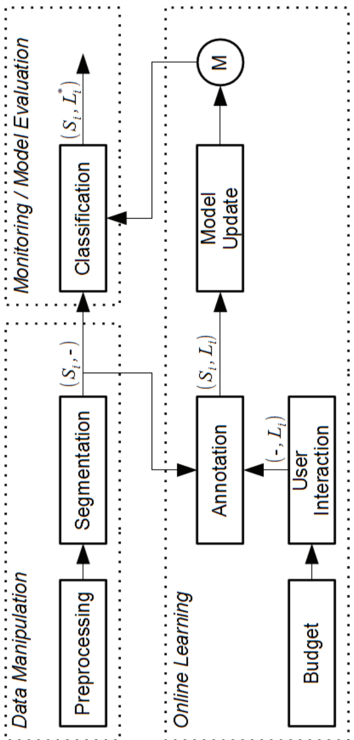

Figure 2: Interactive Pipeline

The issue of how reliably users can be expected to answer questions about past events, limits the applicability of *Active Learning* [19] (AL) for the purpose of acquiring annotations and model learning. AL is based on the idea that a "pool" of un-annotated data points is available to the learner at each step of the learning process. An information gain function is used to select the most informative points in the pool, and requests are issued to the user in order to obtain the corresponding label. This produces an updated pool, which is again evaluated to obtain a new label, iteratively [1]. Unfortunately, the user's memory cannot be trusted for reliable annotation if the target activity took place in the distant past or too many other activities took place after the target activity. This effectively rules out Active Learning techniques for online annotation of activities.

In summary, we argue that current user involvement techniques are ill-suited for online annotation scenarios. In response to this, our work aims at effective ways of bootstrapping HAR systems with users providing ground truth annotation in an economic and acceptable way, leading to the rapid training of functional and extensible recognisers.

A Budget-based Online Annotation Framework

Overview
The focus of our work is on exploring strategies for online annotation of human activities, with emphasis on mobile and ubiquitous scenarios. In such scenarios: *i)* Ground truth annotations are provided by the prospective user of a mobile HAR technology; and *ii)* a budget is available for annotation. Our hypothesis is the existence of a fixed budget which models limited levels of tolerance. The objective of this paper is to explore suitable budget sizes and budget spending strategies.

Experimental evaluations in mobile and ubiquitous computing applications is a challenge in itself as interactive scenarios are difficult to replicate, which poses a challenge to objective judgements and is not appropriate for in-depth exploration. In response to this, we systematically assess the effects of different budget spending strategies, by realistically simulating interactions aimed at selectively acquiring user annotations. This gives us complete control over the selection of the subsets of annotations to use, and provides a level of repeatability which would be very difficult to achieve using a field experiment (in addition to being more practical and economical overall).

Interactive Pipeline
We address a HAR scenario where the system bootstraps a recognizer that is custom-made for each user, by occasionally collecting input from users while they go about their daily living. The system continuously records sensor readings and, according to a schedule, prompts the user to annotate recently identified activities. Because user compliance to interruptions is a limited resource, not everything is annotated, but, rather, a convenient budget and schedule of interruptions is specified in advance.

In order to streamline the interactive bootstrapping process, we propose a data processing pipeline that combines HAR learning procedures from the literature, with the capability to collect user-provided annotations. The interactive pipeline, illustrated in Fig. 2, is generic and can be adapted to the specifics of the HAR application under consideration.

Preprocessing. This step centralizes automatic sensor readings and provides the core machine learning preprocessing functions, such as the definition of a sliding window over which vectors of feature sets are extracted [6].

Segmentation. The preprocessing step produces a sequence of frames. When a frame captures the full characteristics of an entire periodic activity like, for example, walking or running, frames can be used as individual training examples. Non-periodic activities such as those considered here, however, are only fully expressed across multiple frames, suggesting that training examples should consist of sequences of contiguous frames, called *segments* and denoted S_i. It is these segments that the user is asked to annotate and that are used to bootstrap the activity recognizer.

Budget. The decision of when annotation requests should be made to the user is controlled by a budget configuration consisting of a *budget size*, which defines the total number of annotation requests available to the system, and *budget spending strategy*, which determines the distribution of such requests over time.

Interaction with User. This component is triggered by the Budget component and is responsible for obtaining annotations from users in the form of labels L_i. Annotation requests always refer to the most recently identified segment. As mentioned, for the purpose of this paper, user interaction is simulated using ground truth annotations available with the Opportunity dataset.

Model Update. When the training set is extended with a new training example (a segment with an associated label), the system re-trains the model on the newly enlarged training set.

Classification. Since the activity classifier is bootstrapped using incrementally collected activity labels, classification accuracy is expected to increase with the growing size of the training set, as more labels are obtained. Thus, in addition to the final accuracy (corresponding to the point where the entire budget has been spent), in our results we also report the *learning rates*. These are the intermediary classification accuracy scores measured at every stage of the bootstrapping process, namely every time a new training example is supplied and the model is updated.

Annotation. This component acts as a bridge between the machine learning and the interactivity parts of the pipeline. The annotation stage constructs training examples by fusing segments with user-provided activity labels.

Methodology

We use the previously described HAR framework in a realistic case study where we investigate the effectiveness of budget based online annotation in general and in particular focus on the influence of different budget sizes and spending strategies.

Segmentation and Budgeting

Segmentation In our experiments we study the effects of budget configurations on recognition performance. We therefore ignore possible segmentation errors, by assuming, as part of our simulation, a perfect segmentation procedure which identifies the correct boundaries between segments, exactly when there is a change in activity. In our simulations we use the segment boundaries from the ground truth annotations in the dataset. While automatic segmentation is a complex problem in itself, it is nonetheless possible to obtain high recognition accuracy scores [12] so our assumption of ideal segmentation comes close to what is achievable.

Budget Sizes Clearly, the larger the annotation budget, the more segments are available to train a recognizer, which shall result in better recognition performance. Although, realistically, the budget size may be limited by human, context and application considerations, we are interested in studying the relationship between recognition performance and budget. Thus, we experiment with three budget sizes: *small* (10 annotations), *medium* (40) and *large* (100). The choice of budget sizes not only provides insight into expected recognition performance, but also exemplifies how additional annotation effort translates into increased performance.

For comparison, as a reference we use the theoretical best-case scenario where the entire sequence of segments is annotated. This *baseline* provides us with an upper bound in model accuracy.

Budget Spending Strategy The next design choice is how to spend the budget. When a segment is generated, the system makes an online decision whether to interrupt the user to annotate or to discard it. We model the distribution of interruptions over time, using the following four different strategies.

Uniform Random The interruptions are scheduled at random within a horizon of time, according to a uniform probability density function.

Uniform Constant The interruptions are scheduled to occur after fixed time intervals.

Upfront The budget is spent as quickly as possible. For every detected segment, an annotation request is prompted until the budget runs out.

Exponential The density of interruptions is an exponentially decaying function. Interruption times are sampled from an exponential probability density function, so more interruptions are likely to happen at the beginning and very few toward the end of the horizon of time.

Strategies can be chosen such that the budget is expended as soon as possible (**Upfront**), more quickly at the beginning (**Exponential**) or more evenly across time (**Uniform Random** or **Uniform Constant**). Distributing interruptions in time in a certain way is certainly motivated by numerous factors which lie outside our experimental environment. In this paper we are interested in the impact of budget strategies on recognition performance.

In a simulation with a finite dataset, distributing the budget within a fixed horizon of time has the advantage of guaranteeing the spending of the entire budget. In a real-world deployment, similar spending patterns can be alternatively obtained by employing an interruption probability p that would vary with the number of previous interruptions according to the employed strategy.

Evaluation Methodology

We use the popular Opportunity dataset [7] to simulate online bootstrapping of a HAR recognizer using user-provided annotations. We segment the training set into a sequence of segments which serves as input to our incremental learning simulations. User interaction by requests to annotate is simulated by annotating segments using the dataset's ground truth labels (which are provided offline together with the set of sensor readings). We control the annotation patterns by specifying the budget size and strategy.

We measure model accuracy using a separate test set, also segmented, so that testing is done at segment level. We then calculate the model's F-Score with regard to the the segments in the independent test set:

$$F = \sum_i 2w_i \frac{P_i R_i}{P_i + R_i}$$

where P_i and R_i are the precision and recall, respectively, of the classifier on activity a_i. The weighting factor w_i is defined as the relative numerosity of a_i, $w_i = N_i / \sum N_i$, where N_i is the number of segments belonging to a_i in the test set.

Segment shuffling. We report the *learning curves* of the classifier during all stages of the bootstrapping. As learning curves from a single budget expenditure are very jagged, in order to reduce fortuitous performance spikes or drops, we perform 50 repeated randomizations of the activity segments data and then report the average F-scores over all randomizations.

Dataset

Opportunity [7] is a publicly available benchmark dataset widely used in current HAR research. Collected in order to advance the state-of-the-art in terms of HAR, Opportunity contains contiguous sequences of readings from a set of 23 worn sensors by the participants while they perform a predefined set of common gestures or activities of daily living (ADLs).

We use Opportunity to pursue an extensive set of experiments on how to bootstrap recognition systems using online processing techniques. By using Opportunity and by describing our experimental setup, we have ensured that we ground our conclusions in a non-trivial classification task and that our research is reproducible.

Opportunity contains data collected independently for four subjects. Each subject has six data files: `ADL1`, `ADL2`, `ADL3`, `ADL4`, `ADL5` and `Drill`. In our use of the dataset, we follow the gesture recognition task in the challenge definition (Task B2) set out in [7]. As specified, we use the gesture sequences in the subsets `ADL1`, `ADL2`, `ADL3`, and `Drill` as the training set from which we draw activity segments, and the sequences in subsets `ADL4` and `ADL5` as the fixed test set, by which we evaluate the classifier's accuracy at each step of the learning curve. We use a subset of the 23 body-worn sensors available in the files, namely we used signals from 5 tri-axial accelerometers (upper right arm, lower right arm, upper left arm, lower left arm and back).

Each atomic activity, or *gesture segment*, consists of a sequence of adjacent frames annotated with the same activity label, for instance "Open Fridge". In a realistic application scenario we would prompt the human to annotate their activities on this segment-level, i.e., the system would ask for one label per activity instance and then assign the same label to all frames this very segment subsumes.

We follow the suggestion of Rebetez *et al.* [16] who reduced the Opportunity gestures to 7 by aggregating

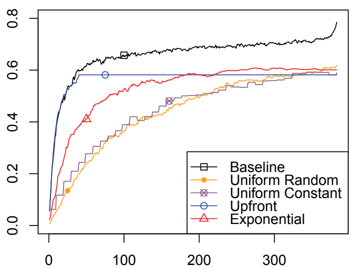

Figure 3: Influence of budget strategy over model bootstrapping speed. X-axis is number of segments **seen**, Y-axis is prediction F-Score.

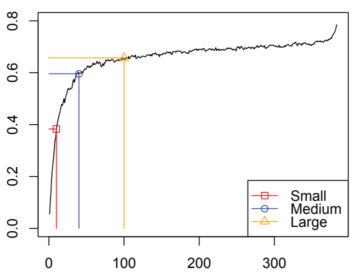

Figure 4: Obtaining performance estimates from baseline graph. X-axis is number of segments **annotated**, Y-axis is prediction F-Score.

similar ones, namely Open/Close_Fridge, Open/Close_Drawer, Open/Close_Door, Clean_Table, Open/Close_Dishwasher, Switch_Light and Drink.

Opportunity contains the *null class* activity label which designates any activity outside the predefined vocabulary of interest – resulting from the aforementioned 17 gestures. We ignore segments labelled as null because the interpretation of these would not generalise well to a realistic deployment.

Classification Backend
Given that the focus of our work is on exploring effective annotation strategies, we employ a standard analysis approach for human activity recognition, which shall be deemed to provide reasonable classification accuracy results [6].

Input data. In this study we focus on tri-axial accelerometer data. Note that this is not a limitation of the presented approach but rather a practical consideration, consistent with the popularity of accelerometry in contemporary HAR applications.

Feature extraction. We employ a standard sliding window procedure (e.g., [15]) that translates the continuous stream of sensor data into a sequence of small analysis frames capturing 500ms of consecutive sensor readings, and overlapping by 50%. For every frame we then calculate the mean of each signal axis – in contrast to more recent feature learning approaches (e.g., [3, 13]) a simple yet reasonable local feature representation (and in line with the Opportunity baseline [17]).

Classification. These feature vectors are then fed into a classification backend, for which we utilise a standard C4.5 decision tree. In doing so we adopt the approach developed by one of the participating, very successful teams in the original Opportunity challenge [7].

Results

Owning to space limitations, we only report the results for Subject 1. These results, however, are representative for the whole Opportunity dataset.

Subject 1 has 383 training set segments. We refrained from analysing null-class segments (as previously explained), as well as from processing segments shorter than the length of a sliding window, and those segments containing missing sensor readings. Evaluation was done against Subject 1's fixed testing set which contains 115 segments.

	Small	**Medium**	**Large**	**Baseline**
UR	0.39	0.62	0.64	0.78
UC	0.39	0.59	0.65	
Upfr	0.39	0.58	0.65	
Exp	0.39	0.60	0.65	

Table 1: Final recognition accuracies (F-scores; Opportunity challenge test set) for different budget configurations.

The experimental results show that it is possible to bootstrap human activity recognizers by involving the user in an online annotation process. Table 1 shows the asymptotic performance that can be expected using our proposed budget configurations. More annotated segments result in better recognition ability.

It is also clear that the budget strategy does not affect the end performance, but it has an effect on the learning rate of the activity model. Figure 3 shows that the strategy impacts the speed with which the recognizer is bootstrapped. We have plotted the performance of all strategies for the Medium budget size (40 units) and also for the baseline. The x-axis is the number of processed or seen activity segments. As explained before, if the budget size is strictly less than the total number of segments, not all processed segments are annotated. Figure 3 shows that strategies that request annotations early on, such as **Upfront** or **Exponential** cause a steeper learning rate - they reach the end performance level sooner. In contrast, lazier strategies such as **Uniform Random** or **Uniform Constant** delay the production of a reliable activity model but do not amass interruptions early on.

We have isolated the learning curve of the baseline illustrated in Figure 3 and displayed it in Figure 4. This shows an exhaustive analysis of budget sizes, where the x-axis represents the budget size and the y-axis is the expected *end* recognition accuracy. The budget strategy, as we have seen, determines how fast the end performance is going to be reached. This can be used to inform budget sizes of real user deployments.

Summary and discussion

Learning accurate Human Activity Recognition models requires training examples which are often difficult to acquire in practice. Our work is set in the context of online learning, where further challenges arise. Firstly, labelled examples only become available incrementally, as the activities unfold. Secondly, labels must be acquired through proactive interaction with the user, who may have limited tolerance for such interruptions, as well as limited memory to recall past events. This leads to the notion of a *budget* of available user interactions, whereby the user is asked to identify the type of activity associated with the most recent gesture. The combination of these factors leads to a scenario where the learning process can only afford a set number of interactions, which are aimed at labelling the type of activity that is being observed.

Summary of contributions

In this paper we proposed a principled way of analysing the trade-offs between the number of available interactions (budget), the way the budget is spent over time (budget spending strategy), and the accuracy of the HAR models that can be learned under such budget constraints. Our approach involves replaying segments from the Opportunity challenge dataset and simulating interactions that occur during sequences of activities for an extensive set of budget configurations.

Our main contribution is an experimental method which is generally applicable to the online learning setting. Our results indicate that (i) recognition accuracy close to the baseline can be achieved by using about 50% of the labels that are potentially available; (ii) the choice of budget spending strategy has little bearing on overall accuracy at the end of training, however it affects the learning rate, which certainly has massive implications on the overall acceptability of user-involvement in online learning of HAR systems.

Open questions concern the determination of realistic budget sizes and spending strategies. We expect our results to be instrumental to inform future user studies, including our owns.

Acknowledgements

The authors would like to thank Marianna Obrist and James Turland for their valuable suggestions. Parts of this work have been funded by the RCUK Research Hub on Social Inclusion through the Digital Economy.

References

[1] Alemdar, H., Kasteren, T. v., and Ersoy, C. Using Active Learning to Allow Activity Recognition on a Large Scale. In *Proc. AmI* (2011).

[2] Bao, L., and Intille, S. S. Activity Recognition from User-Annotated Acceleration Data. In *Proc. Pervasive* (2004).

[3] Bhattacharya, S., Nurmi, P., Hammerla, N., and Plötz, T. Using unlabeled data in a sparse-coding framework for human activity recognition. *Pervasive and Mobile Computing* (2014). in press.

[4] Brand, A. Public health genomics and personalized healthcare: a pipeline from cell to society. *Drug Metabolism and Interactions* (2012).

[5] Bull, F. C., Kreuter, M. W., and Scharff, D. P. Effects of tailored, personalized and general health messages on physical activity. *Patient Education and Counseling* (1999).

[6] Bulling, A., Blanke, U., and Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys* (2013).

[7] Chavarriaga, R., et al. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters* (2013).

[8] Cohn, D. A., Ghahramani, Z., and Jordan, M. I. Active Learning with Statistical Models. *Journal of Artificial Intelligence Research* (1996).

[9] Huynh, T., Fritz, M., and Schiele, B. Discovery of activity patterns using topic models. In *Proc. UbiComp* (2008).

[10] Intille, S. S., Bao, L., Tapia, E. M., and Rondoni, J. Acquiring in situ training data for context-aware ubiquitous computing applications. In *Proc. CHI* (2004).

[11] Lim, B. Y., and Dey, A. K. Investigating Intelligibility for Uncertain Context-Aware Applications. *Proc. UbiComp* (2011).

[12] Okeyo, G., Chen, L., Wang, H., and Sterritt, R. Dynamic sensor data segmentation for real-time knowledge-driven activity recognition. *PMC* (2014).

[13] Plötz, T., Hammerla, N., and Olivier, P. Feature Learning for Activity Recognition in Ubiquitous Computing. In *Proc. Int. Joint Conf. on Art. Intelligence (IJCAI)* (2011).

[14] Plötz, T., Hammerla, N., Rozga, A., Reavis, A., Call, N., and Abowd, G. D. Automatic Assessment of Problem Behavior in Individuals with Developmental Disabilities. In *Proc. UbiComp* (2012).

[15] Plötz, T., Moynihan, P., Pham, C., and Olivier, P. Activity Recognition and Healthier Food Preparation. In *Activity Recognition in Pervasive Intelligent Environments*. Atlantis Press, 2010.

[16] Rebetez, J., Satizábal, H. F., and Perez-uribe, A. Reducing User Intervention in Incremental Activity Recognition for Assistive Technologies. In *Proc. ISWC* (2013).

[17] Sagha, H., Digumarti, S. T., Chavarriaga, R., Calatroni, A., Roggen, D., and Troester, G. Benchmarking classification techniques using the Opportunity human activity dataset. In *Proc. SMC* (2011).

[18] Scollon, C. N., Kim-Prieto, C., and Scollon, C. N. Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *Journal of Happiness Studies* (2003).

[19] Settles, B. Active learning literature survey. *University of Wisconsin, Madison* (2010).

[20] Stikic, M., Larlus, D., Ebert, S., and Schiele, B. Weakly Supervised Recognition of Daily Life Activities with Wearable Sensors. *TPAMI* (2011).