The Probabilistic Activity Toolkit: Towards Enabling Activity-Aware Computer Interfaces

Matthai Philipose[1], Kenneth P. Fishkin[1], Mike Perkowitz[2], Donald Patterson[1,2], Dirk Hähnel[1,3]

Intel Research Seattle, Seattle, WA 98115, USA
Dept. of Computer Science and Engineering, University of Washington, Seattle, WA 98115, USA
Dept. of Computer Science, University of Freiburg, Freiburg, Germany
{ matthai.philipose, kenneth.p.fishkin, }@intel.com, {map, djp3}@cs.washington.edu,
haehnel@informatik.uni-freiburg.de

IRS-TR-03-013

November 2003

Intel Research Seattle Tech Memo IRS-TR-03-013, November 2003

The Probabilistic Activity Toolkit: Towards Enabling Activity-Aware Computer Interfaces

Matthai Philipose[1], Kenneth P. Fishkin[1], Mike Perkowitz[2], Donald Patterson[1,2], Dirk Hähnel[1,3]

Intel Research Seattle, Seattle, WA 98115, USA

Dept. of Computer Science and Engineering, University of Washington, Seattle, WA 98115, USA

Dept. of Computer Science, University of Freiburg, Freiburg, Germany

{ matthai.philipose, kenneth.p.fishkin, }@intel.com, {map, djp3}@cs.washington.edu, haehnel@informatik.uni-freiburg.de

ABSTRACT

Emerging HCI techniques require the ability to recognize activities that occur in the physical world. Systems that recognize home activities have been limited in the variety of activities they recognize, their robustness to noise, and their ease of use. We present a toolkit (PROACT) for activity recognition that addresses these problems by leveraging three novel techniques: automatically mining text documents and the web for activity structure; recognizing object use via Radio Frequency Identification (RFID) technology; and combining these two inputs to infer user behavior with a flexible and scalable, Monte-Carlo based inference engine. As an initial evaluation, we successfully applied our system to a known difficult problem from health care: recognizing multiple "Activities of Daily Living" (ADLs) in a real home environment. Promising results from a user study validate PROACT's approach.

Keywords

Ubiquitous, proactive or context aware computing, RFID, activity inference, data mining, machine learning, ADLs

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. J.3 Life and Medical Sciences: health

INTRODUCTION

Many HCI researchers are pursuing visions of "invisible", "disappearing", "proactive", or "polite" human-computer interaction [cf. 5, 10, 12, 26]. In this paradigm, computers observe the behavior of users and their environments via networks of sensors, and perform designated tasks on behalf of the user, ideally with little explicit user direction.

At the core of these systems is a module that uses sensor data to infer aspects of the users' current *context.* An important part of context is the activities currently being performed by the users, and the amount of progress that they have made toward completion. For example, it is reasonable to believe that a context-based reminder program would be more effective in reminding people to balance their checkbooks if it did so while they were paying bills rather than when they passed through the study to answer a phone call.

Engineering such modules to recognize human activity, particularly in the home, is a challenge. Most applications approximate activity recognition with proxies such as room location and identity [3, 7, 8]. Systems that do finer-grained recognition are rare, and typically limited to a specific activity. They are painstakingly engineered using custom sensors, heuristic reasoning engines and rooms-of-the-future [2, 15, 16, 27]. We present here a probabilistic activity toolkit, PROACT, that can recognize routine activities in structurally un-modified rooms with reasonable accuracy. PROACT requires one simple, robust, easy to deploy sensor. We believe that this toolkit can be useful in making context-aware applications more intelligent, and we demonstrate this via a preliminary evaluation.

For our evaluation, we applied PROACT to the problem of reporting 14 Activities of Daily Living (ADLs) [13, 14], a standard metric used to measure the cognitive wellness of people requiring assisted care. ADLs are a particularly interesting domain of application because the activities to be monitored are standardized, matched to a real-world need, and known to be tedious and error-prone to monitor. Accordingly, automatic ADL inference has been intensively investigated [2,3,7,8,15,16,21,27,28] with limited success. We instrumented a real home and tested our system in that home over a 6 week period with 14 non-researcher users.

The early results are promising. We show that nine ADLs, which no known prior work has addressed, were accurately inferred by PROACT. For four more ADLs, we were able for the first time to move beyond qualitative presentation to quantitative analysis.

At the heart of PROACT is a breakthrough in sensing technology. Recent advances in miniaturization and manufacturing have dramatically improved the functionality and reduced the cost of transceivers, called Radio Frequency Identification (RFID) tags, which can be attached unobtrusively onto individual objects as small as a plastic fork. These tags are the size of a postage-stamp, require no batteries and are inexpensive. When interrogated by a radio they respond with a globally unique identifier. Combined with PROACT's tag readers, this technology reliably identifies objects that are touched, moved, or co-located by a person.

The ability to detect a large number of objects relevant to an activity has novel implications for activity definition and detection. In particular, we define an activity in terms of the probability and sequence of specific object involvement (touching, movement, co-location). "Paying Bills", for instance, may involve bills, folders, calculators, check books, pens and computers. The structure of our models allows us to define new activities only by converting text definitions of activities (structured as "recipes" or how-to's) with lightweight English text-processing. In fact, we have mined the web for roughly 20,000 activity models so far. Finally, since many activities are quite strongly characterized by the objects involved, simple models suffice. By avoiding intricate causal structure for activities while associating as many relevant objects as possible, our models are both fast and precise.

In what follows, we first describe how PROACT is intended to be used and explain its design. Next, we describe our evaluation, including further details on the application domain (ADLs), the experiment, and the results. Finally, we discuss related work before concluding and presenting future work.

USAGE MODEL

PROACT assumes that "interesting" objects in the environment contain *RFID tags*. These can be purchased off the shelf, cost roughly $0.40 each, have the form factor of postage stamps (including adhesive backing), and can withstand day-to-day use for years. PROACT deployment involves tagging tens to hundreds of objects in their environment. This can be done incrementally; the more tags, the broader and deeper the inferencing. Tagging an object involves sticking an RFID tag on it, and making a database entry mapping the tag ID to a name. Despite this apparent overhead, market forces will soon automate this infrastructure creation [1].

Users then employ *RFID tag readers* to monitor their activities. They may wear tag-detecting bracelets or gloves, place long-range (roughly 20ft) readers in corners of rooms, or run robots, vacuum cleaners, or janitorial carts, with mounted long-range readers. As users go about their daily activities, the readers detect tags that (a) users touch, (b) are close to them, or (c) are moved by them, and thereby indirectly deduce which objects are currently "involved" in their activity. PROACT uses the sequence and timing of object involvement to deduce what activity is happening.

PROACT is intended to run in real time. An application can query it at any time for the likelihood of various activities being tracked or details of those activities (e.g., objects involved, or durations), or subscribe for event notification when activities occur with a specified degree of certainty.

Programmers name activities using plain English phrases (e.g. "paying bills"). The phrase used can either be chosen from a list provided by PROACT, or can be a new one provided by the programmer. For a new phrase, the programmer can define the activity by providing a text document containing an English description of the steps involved in the activity (much like a recipe), or request that PROACT mine the definition automatically from the web. In either case the mining engine converts text into activity definitions.

SYSTEM OVERVIEW

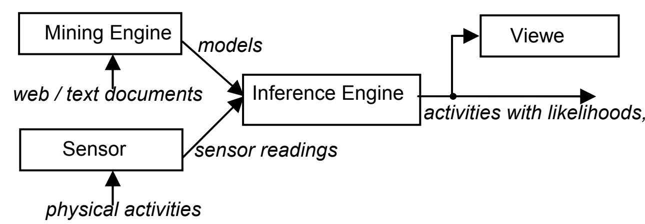

Figure 1: A High-Level View of PROACT

Figure 1 presents the main components of PROACT. It is centered on an inference engine which, given models for activities, and sequences of sensor readings, returns the likelihood of current activities. The models are produced by the mining engine, which extracts them automatically from text documents, including but not limited to websites. The sensor readings are produced while the end-user performs activities. For debugging, PROACT provides an activity viewer, which provides programmers with a real-time view of activities in progress, the sensor data seen, and how belief in each activity changes with the data.

Activity Model Design

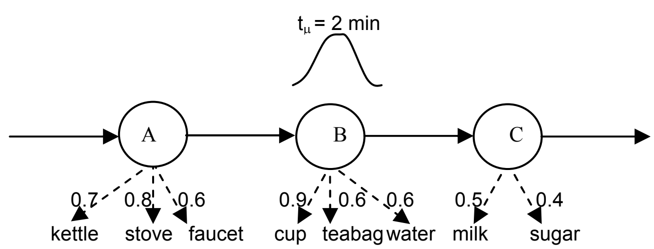

Figure 2: PROACT Model for Making Tea

PROACT's activity model is restricted to linear sequences of sub-activities which are annotated with timing and object information. Figure 2 gives an example of how to model the activity "making tea." The activity consists of three consecutive sub-activities, drawn as circles (corresponding to ()A)boiling water, (B)steeping, and (C)flavoring the tea).

PROACT allows each sub-activity to have its time-to-completion be modeled as a Gaussian probability distribution. In this example, steeping the tea is expected to take 2 minutes. The amount of time required to boil the water and flavor the tea is unknown, and does not influence the reasoning.

The final part of the representation is a dotted arrow denoting the probability that an object is involved in an activity. In Figure 2, for instance, we expect to see a kettle 70% of the time that we are boiling water, whereas sugar is involved in the mixing phase 40% of the time. This probability combines three sources of ambiguity: sensor error, model error, and modeling generality (e.g., water may be in the kettle, and as a result the faucet may not be touched).

Mining Activity Models from Text Definitions

Since we expect that it will be difficult for lay programmers to directly model activities in this fashion, PROACT allows activities to be specified as text documents that are structurally very similar to recipes.

Making Tea:

1. Fill a **kettle** from the **faucet**. Place kettle on the **stove** and boil.
2. Pour hot **water** into a **cup**, filling ¾ of the cup. Immerse **teabag** in cup for two minutes and dispose of teabag.
3. Add **milk** and **sugar** to taste.

Figure 3: Steps for Making Tea

Figure 3 provides an example. Each document has a title, an optional list of objects involved, and a step-by-step description of how to perform the activity. The mining engine then converts the document into an activity model by interpreting the steps as sub-activities and the objects mentioned in each step as the set of objects involved. When available, timing information is taken from the description.

PROACT does not require full natural-language processing: light-weight syntactic analysis suffices to identify the steps. To identify the objects mentioned in each step (highlighted in bold in the figure), we use WordNet [28], a package which maps words to their potential uses. Our activity description requirements are simple enough that by adding a custom HTML parser, we are able to successfully mine recipe and how-to sites. PROACT comes with roughly 20,000 predefined activity definitions including "cleaning a bathtub" and "boiling pasta".

PROACT determines the object involvement probabilities p in a novel manner. The method relies on a "Mirror Assumption": if an activity name a_n co-occurs with some object name o_n in human discourse, then activity a is likely to involve object o in the physical world. We postulated on this basis that if a_n occurs on n_1 pages on the web (which we treat as a compendium of human discourse), and there are n_2 pages containing both a_n and o_n, then $Pr(o|a) \approx n_2/n_1$. We obtain these numbers via the Google programming API. There are clearly instances when this assumption fails; however, our experience has been that it is sufficient to generate an initial set of probabilities: they were used, unchanged, in our experiment.

Sensing objects of interest via RFID

Figure 4: RFID Readers: Robot (L) and Glove(R)

Activity inference depends on being able to observe objects that are "involved" in activities. PROACT currently uses RFID readers in two ways to determine two types of involvement. First, it uses long-range readers mounted on a mobile robot platform to map the location of objects in the activity space, and coupled with the location of the user, the set of objects near a person at any given time. Second, it uses a short-range reader built into the palm of a glove that can determine the objects that are touched. Both readers are built from commercial off-the-shelf equipment.

The antenna built into the palm of the prototype glove is connected to an Intermec RFID reader, which is packaged with a Crossbow Mica Mote radio, a USB-based power supply, and a rechargeable battery. All components except the antenna are housed in the small box on the glove of Figure 4. The reader samples the environment once every half second; any RFID seen is broadcast to an HP iPaq 5400, utilized as a wearable computer. The iPaq either stores the data onboard for future analysis or forwards it via WiFi to the inference engine running on a workstation. The reader lasts for two hours at this duty cycle. PROACT presently supports two mobile RFID readers – one robot-based, and one glove-based, as shown in Figure 4. For this experiment, we used the glove.

While it would be problematic for many applications to use such a glove, this is a temporary problem. When a portable RFID glove as a UI device was proposed 3 years ago [25], it measured roughly 160 cm^3, and was not wireless. Our latest version is a little over 30 cm^3, including wireless. Within a few years portable RFID readers can fit into a large wristwatch or bracelet, and can detect RFID tags near the palm of the hand.

The RFID-reading robot, which has two RFID-reading antennas and a laser scanner onboard, has to overcome two challenges. It has to localize itself relative to the space it scans, and it needs to localize tags relative to itself. Combining the two allows it to provide absolute locations for the tags. We use a laser scanner to achieve the former, using recent advances in Simultaneous Localization and Mapping [20] from robotics to combine scanner readings that build a map and localize a robot, even in completely new spaces. Using statistical techniques, we can take the 3 to 6m detection ranges and localize tags to within 1m.

Activity Inference

By regarding the current sub-activity as a hidden variable, and objects seen and time elapsed as observed variables, the problem of probabilistically estimating the hidden variables given observables is handled by the general technique of Bayes Filtering. Our activity inference engine converts the activity models produced by the mining engine into Dynamic Bayesian Networks. We use a type of Sequential Monte Carlo (SMC) approximation called "particle filters" to probabilistically solve for the most likely activities [6]. Our work is adapted from related work on transportation behavior inference [23]. Intuitively, activity inference may be understood as follows. Initially, the inference engine picks a large number of "particles" and distributes them uniformly over all the sub-activities. At any given time during the inference, the probability that the system is in a particular sub-activity is equal to the fraction of particles in that sub-activity. Each time the system sees an observation (including clock ticks), the number of particles in sub-activities which are supported by the observation is increased, and the number in other sub-activities is correspondingly decreased. In this way, all activities are reasoned about in parallel.

EVALUATION

The primary goal of our evaluation was to test whether PROACT could infer the correct activities from a useful set performed by people in a realistic setting. Below, we first discuss the activity domain for the experiment, and then provide details on the experiment and its results.

The Domain: Activities of Daily Living (ADLs)

We evaluated PROACT with a known difficult set of problems in activity inferencing: inferring daily activities in the home. The health care community refers to two sets of daily activities; those which involve interactions with "instruments" (telephones, appliances, etc.), termed Instrumental Activities of Daily Living (IADLs) [14], and those which do not ("ADLs") [13]. We follow standard practice and use "ADL" generally to refer to both sets. ADLs have become a standard set of activities to infer for two reasons:

First, ADL monitoring is an ongoing, important activity in health care. For example, in the United States, any nursing home that receives Medicare funds has to record and report ADLs. Trained caregivers spend a great deal of time measuring and tracking ADL accomplishment for persons under their care. However, this monitoring is difficult. It is time-consuming, prone to forgetting (both forgetting that an ADL was observed and forgetting to record it), and invasive. Automated aids that can address these issues and augment caregiver work practice are of great interest. We stress that PROACT is not designed to replace a caregiver, but rather to augment and ease their work practice. Robust and potentially unseen RFID tags are particularly well-suited for ADL monitoring of people with early-stage dementia (an increasingly common and problematic target population), as they are prone to break and disassemble fragile and/or unusual objects.

Second, ADLs are general and challenging because they are common activities that people engage in daily. These activities are useful to infer in the home setting as well, but this is particularly challenging because the home is a private, difficult to instrument, and fluid setting.

Of the more than 20 ADLs that caregivers can monitor, we chose 14 for evaluation. We eliminated the rest due to the structure of our experiment, where subjects performed the tasks in another person's home, so ADLs such as "bathing" could not feasibly be tested. Others were omitted because they required travel outside the home (e.g. "shopping", "crossing the street"). We stress that it was the focused nature of the experiment, not any limitation of the toolkit, which led to this exclusion. The 14 ADLs we tested (shown in Column 2 of Table 1) are, to our knowledge, 11 more than any other system has attempted.

No.	ADL	Task Sheet Description
1	Personal Appearance	Please touch up your personal appearance.
2	Oral Hygiene	Please take care of your oral hygiene as if you are about to go to bed.
3	Toileting	Please use the toilet.
4	Washing	Please wash up.
5	Housework	Please clean something.
6	Safe use of Appliances	Please use an appliance.
7	Use of heating	[1]Please adjust the thermostat.
8	Care of clothes and linen	Please use the washer and/or dryer.
9	Preparing simple snack	Please prepare something to eat.
10	Preparing simple drink	Please make yourself a drink.
11	Use of telephone	[2]Please use the Telephone to get your horoscope.
12	Leisure Activity	Please watch TV, read, play cards, or listen to music for a few minutes.
13	Caring for an infant	[3]Please care for the baby.
14	Taking Medication	[4]Please take some pills.

Table 1: tested ADLs, and their descriptions

Notes:

1 "Use of heating": the thermostat was the only heating control in the house.

2 "Use of telephone": we focused the task to encourage a short call.

3 "Caring for an infant": a life-size doll was used, but not tagged.

4 "Taking Medication": the pills were replaced with candy.

Experiment

Experimental Procedure

We felt it vital to use a real house, with real objects, in real locations, used by real people and subject to the real wear and clutter of daily living. In a perfect world, we would have instrumented the houses of each of our subjects, but this was impractical. Instead, we chose one inhabited house (inhabited by one of the authors, his wife, and their 2-year-old child) and instrumented it with 108 tags in a few hours. This approach has disadvantages. Subjects weren't in their own homes, and we could not do longitudinal study. However, we feel this was a valid compromise. Multiple subjects volunteered to us that since they were not observed and were in a real house, they felt relaxed and at ease, and that they acted naturally.

Figure 5 shows the kitchen area with some tags circled. We did this before the activity models existed, so as to not be biased by its later designation of "key" objects. By tagging as many objects as possible, we also hoped to avoid having subjects feel they were "steered" to a narrow set of objects.

Figure 5: Kitchen of Experiment Home

Over the next 6 weeks, we had 14 subjects (3 M, 11 F) perform their activities, unobserved, in this house, wearing the glove of Figure 4. Each spent roughly 45 minutes in the house. Ages ranged from 25 to 63, with a mean of 39.

After a tour of the house, and a demonstration of the RFID glove, subjects were informed of the nature of the experiment. Subjects were given a package of 14 task sheets, one for each ADL. Table 1 shows the tasks as described on each sheet. The task sheets also had pictures of where some objects could be found in the house to avoid "treasure hunts". As Table 1 shows, we kept activity descriptions as broad and as close to the medical description as possible.

Subjects randomly selected 12 of the 14 tasks. They then went into the house, and performed those 12 tasks, in any order, without observation. Subjects could engage in other household activities as they wished. As they touched tags in the course of performing their tasks, they would hear a "beep", as PROACT indicated that it had recorded a tag touch. Since this round of experiments was not meant to test the efficacy of the glove itself, subjects were asked to touch something several times if they saw that they were touching a tagged object and didn't hear a "beep". This was sometimes necessary due to the prototypical nature of the glove; typically, subjects touched around 60 sixty tags, one of which needed to be repeated. While performing tasks, subjects wrote on a sheet of paper which task they were doing. After leaving the house, subjects gave one experimenter the sheet. This was kept separate from the other experimenters until after the results were processed. Subjects were paid $20.

As an additional test of the robustness of the system, the tags were left in the house while subjects were not present. In total, the tags were in place for 6 weeks in the house, which was permanently occupied and used by an extended family (the house was also frequently lived in by visiting relatives). With the exception of a few tags that were awkwardly placed on edges all tags stayed usable throughout the entire experiment, surviving the full destructive power of a 2-year-old in his native habitat.

Results

We gave the complete tag sequence for each subject (without manually segmenting between activities), along with models for the 14 activities to the inference engine. The engine returned a log of the most likely sequence of activities that would explain the readings. Running the activity inference took less time than performing them. Thus, although we used PROACT offline, it is capable of inferring these ADLs in real time.

Recall that PROACT goes beyond a simple 1-bit notification that an activity occurred, and can provide much more detail. For example, for the "preparing a simple snack" task, PROACT can report how long it took to make the snack, which ingredients were used, which utensils were employed, etc. By monitoring these details (are they taking longer to make their meals? Are they wearing the same clothes each day?) , caregivers can get a richer, nuanced view of condition. However, it is difficult to quantify this additional information, so for the analysis below we restrict ourselves to activity-reporting as simple binaries.

We compared the logs generated by PROACT with the activity sequence reported by the subjects immediately after the experiment. We treat the latter as ground truth. Each time the logs claimed an activity occurred, if it was correct that was a true positive (TP), if it wasn't it was a false positive (FP). If a true activity occurred and wasn't reported by PROACT, it was scored as a false negative (FN). Column 2 of Table 2 shows these numbers.

ADL No.	TP	FP	FN	Prec %	Rec %
1	11	1	1	92	92
2	7	3	2	70	78
3	8	3	3	73	73
4	3	0	6	100	33
5	8	0	4	100	75
6[1]	21	4	6	84	78
7	8	0	3	100	73
8	7	0	2	100	78
9	6	2	4	75	60
10	9	5	5	64	64
11	11	0	3	100	79
12	7	0	5	100	58
13	13	0	1	100	93
14	9	0	2	100	82
TOTAL	128	18	47	**88**	**73**

Table 2: Results of Experiment

We then use the standard metrics of precision and recall to summarize the effectiveness of PROACT. *Precision* for an activity is the probability that a given inference about that activity is correct; TP/(TP+FP). *Recall* is the probability that a given true activity will be inferred correctly; TP/(TP+FN). Precision and recall are termed "positive predictive value" and "sensitivity", respectively, in the medical community.

Discussion

As Table 2 shows, PROACT did well on average: when it inferred an activity, it was correct 88% of the time. For many activities, it never incorrectly inferred occurrence of the activities. Of the activity instances that actually happened, it detected 73% correctly. Given the ambiguous and overlapping activity definitions (consider "personal appearance", "oral hygiene", and "washing up" – we did nothing to help PROACT make these subtle distinctions) PROACT did quite well.

To place these numbers in further context, note that while ADL inferencing has often been investigated, 9 of the ADLs inferred here are inferred for the first time. For 4 more (meal preparation, toileting, heating control, and medication taking), this is the first time that any quantitative results are reported. For the one ADL where previous quantitative results were presented, hand washing, PROACT's precision/recall of 100/33% are below the 95%/84% reported by Mihailidis [16], however that work targets exactly one activity and requires cameras to be installed in the bathroom.

The performance on hand-washing is actually our worst case. We now discuss the reasons for this result, and a few others of interest. The radio waves used by most RFID tags are absorbed by water and metal; metal can also short-out the tag antenna. These factors cause the detection rate to plummet for tags which are too close to those substances. The tags on faucets, soap bottle and refrigerator handle were especially affected by this. Activities where touching faucet and soap or the refrigerator were key, namely washing hands (activity 4), making a snack (9) or drink (10) were directly affected by this. More careful tag positioning, and/or using newer RF tags which are optimized for placement near liquids and metal, could mitigate this.

In some cases, the model involves so few observations as to not be easily distinguishable from noise. For example, the only observable for adjusting heat (7) was touching the thermostat; that for playing music (11) was the stereo (CDs were not tagged). The single observation that results is not enough to convince the activity inference engine that a new task has begun. Adding more relevant tags could solve this (e.g. tagging the CDs).

Activities with common prefixes posed a more subtle problem. For instance, activities for personal appearance (activity 1), oral hygiene (2), toileting (3) and washing (4) all begin with entering the bathroom and possibly turning on the light. Our models replicate the nodes for these sub-activities for each of the four activities. When bathroom and light are detected, each of these activities are equally likely (nominally 25%). When disambiguating objects are then seen (e.g. toothbrush and toothpaste), the inference engine concludes that the oral hygiene activity is the correct one. If the user then uses the toilet without leaving the bathroom first, the inference engine fails to detect the second activity. To solve this problem, it may be necessary to consider more nuanced representations of activities.

Although it is possible for PROACT to learn parameters without supervision (in particular, duration of sub-activities) in the style of [23] our experiments did not include any model learning. We used a fixed small number (20 sec) for mean duration, which was sometimes inappropriate. A particularly insidious effect surfaced when

[1] The "Use of an appliance" task is also an optional sub-task in many others: using the microwave to boil water for a drink, using the dryer to care for clothes, etc. We count use of an appliance during another task as a "positive" for the "use an appliance" task.

subjects interleaved boiling water for tea or coffee (10) with other activities. Since boiling takes several minutes, PROACT would conclude that they were no longer boiling water after a few seconds (yielding false negatives), and erroneously "jump" to other activities for the post-boiling events (yielding false positives). We can most likely achieve much better tuning of the time parameter via use of unlabelled training data, but more sophistication will be required of our inference engine to detect interleaved activities.

RELATED WORK

Sensor streams feeding Bayesian networks for activity inferencing have been employed with great success to analyze the behavior of office workers [9, 10, 22] and cell-phone users [16]. However, the qualitatively different nature of the task domain and setting and the sensors available (keyboards, mice, cameras, microphones, cell-phones, etc.) has not yet allowed these techniques to be employed in the sensor-impoverished, unstructured home environment inhabited by ADLs.

Our activity inference engine builds primarily on work on robot GPS location tracking with noisy sensor data. [17, 23]. It is related to work on care-giving with robotic nurses, but does not require manual input of activity completion and can re-estimate model parameters in an unsupervised manner [4,18,24].

Many researchers have investigated ADL inferencing. These investigations fall into two categories: investigations into general solutions for multiple ADLs, and investigations into solutions for single ADLs.

Single-ADL investigations have focused on hand washing [16], medication-taking [28], meal preparation [2, 27], general activity level [15], and use of heating [21]. Due to the low-level information provided by the sensors they employ, and the lack of leverage between investigations due to their idiosyncratic sensor and algorithm suites, only one ([16]) reports the quantitative results of user testing.

Multiple-ADL inferencing has used a variety of low-level sensors (mainly motion and contact) to infer general activity level, and used motion in a room to roughly estimate kitchen and bathroom activity [7, 8, 3]. One [7] reports anecdotal "reasonable assurance" that their inferences work when tested on a single user, but regretfully the inferencing is quite coarse, reporting mainly room-level motion.

The multiple-ADL approach most similar to ours is the current work of Intille et al. [11]. They also use adhesive sensors which are placed throughout the home. Sensor readings are then later analyzed for a variety of applications – kitchen design, context inferencing, and eventually ADL monitoring. They share our contention that robust, easily deployed, cheap sensors measuring real activity in real homes are vital for true user interaction and understanding. Our work differs from theirs in four ways – infrastructure, approach, analysis, and experience. For infrastructure, we use commercially available RFID tags, whereas they employ custom-made sensors. RFID tags are roughly 100 times cheaper, more easily deployable, less physically obtrusive (they can even be unseen), and more physically robust. From an approach point of view, their sensors passively collect data which is then analyzed weeks or months later, after the sensors are removed, while our system is designed to support immediate analysis, with the sensors remaining *in situ*. From an analysis point of view, their work is in its early stages, and is focused on how to best collect and integrate the data, with analysis only beginning. We report a functioning general framework which performs this analysis. Finally, from an experience point of view, we are reporting for the first time an experiment to test and validate such analysis.

CONCLUSION

We believe that PROACTs combination of recent advances in three areas (RFID, data mining, and machine learning) offers great promise for future work in activity inferencing. As shown here, we have successfully employed it to a known problem (ADL monitoring) and achieved good first results in detecting many activities.

There are many opportunities for future work, in each of these three areas. In the sensor domain, we would like to move beyond the glove, and to integrate other types of sensors, particularly location sensors. In data mining, we are investigating how to mine better models, and to suggest "eigenobjects" to be tagged for a given activity set. Key challenges in activity inference include modeling interleaved activities, and those involving many people. Finally, we would like to demonstrate the breadth of PROACT by applying it to a non-ADL domain.

ACKNOWLEDGMENTS

We thank our subjects for their time and feedback, the Intel Proactive Health SRP for suggesting ADL monitoring as a task domain, Henry Kautz and Dieter Fox of Washington CSE for help on the design of the tracker/learner, and the family members of the augmented house for opening their house to strangers for 6 weeks.

REFERENCES

1. Auto-ID center. www.autoidcenter.org/ aboutthetech_idiotsguide6.asp
2. T. Barger, M. Alwan, S. Kell, B. Turner, S. Wood, and A. Naidu, "Objective remote assessment of activities of daily living: Analysis of meal preparation patterns," Medical Automation Research Center, University of Virginia Health System, Poster presentation, 2002.
3. E. Campo and M. Chan, "Detecting Abnormal Behavior by Real-time Monitoring of Patients". AAAI workshop on "Automation as Caregiver", 2002, pp. 8-12.

4. D. Colbry, B. Peintner, and M. Pollack. "Execution Monitoring with Quantitative Temporal Bayesian Networks." In 6^{th} *International Conference on AI Planning and Scheduling*. April 2002.

5. The Disappearing Computer Initiative. www. disappearing-computer.net/

6. A. Doucet, N. de Freitas, and N. Gordon. Sequential Monte Carlo Methods in Practice. Springer-Verlag, Berlin, 2001

7. A. Glascock, D. Kutzik, "Behavioral Telemedicine: A New Approach to the Continuous Nonintrusive Monitoring of Activities of Daily Living". Telemedicine Journal, 6(1), 2000. pp. 33-44.

8. V. Guralnik and K. Haigh, "Learning Models of Human Behaviour with Sequential Patterns". AAAI Workshop on "Automation as Caregiver", 2002, p. 24-30

9. E. Horvitz, A. Jacobs, D. Hovel. "Attention-Sensitive Alerting". UAI '99, pp. 305-313.

10. S. Hudson, J. Fogarty, C. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J.C. Lee, and J. Yang, (2003). Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. CHI 2003. Pp. 257-264.

11. S. Intille, E. Munguia Tapia, J. Rondoni, J. Beaudin, C. Kukla, S. Agarwal, L. Bao, and K. Larson, "Tools for studying behavior and technology in natural settings," *UBICOMP 2003* (to appear).

12. Invisible computing. invisiblecomputing.org/scope.html

13. S. Katz, A. Ford, R. Moskowitz, et al: Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychological function. JAMA 1963;185:914-919.

14. S. Katz. Assessing Self-Maintenance: Activities of Daily Living, Mobility and Instrumental Activities of Daily Living. Journal of the American Geriatrics Society, 1983, 31(12); 721-726.

15. I. Korhonen, P. Paavilainen, A. Sarela. "Application of ubiquitous computing technologies for support of independent living of the elderly in real life settings", Ubicomp 2003 UbiHealth workshop (to appear).

16. P. Korpipää, M. Koskinen, J. Peltola, S-M. Mäkelä, and T. Seppänen. "Bayesian approach to sensor-based context awareness". Pers. Ubiquit. Comput (2003) 7:113-124.

17. L. Liao, D. Fox, J. Hightower, H. Kautz, D. Schulz. "Voronoi Tracking: Location Estimation using Sparse and Noisy Sensor Data." In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. (2003)

18. C.E. McCarthy and M.E. Pollack. "A Plan-Based Personalized Cognitive Orthotic." In 6^{th} *International Conference on AI Planning and Scheduling*. April 2002

19. A. Mihailidis, G. Fernie, J.C. Barbenel (2001). The use of artificial intelligence in the design of an intelligent cognitive orthosis for people with dementia, Assistive Technology., 13: 23-39.

20. M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. *FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem*. AAAI 2002, pp. 593-598.

21. M. Mozer, "The Neural Network House: an environment that adapts to its inhabitants,". *AAAI Spring Symposium on Intelligent Environments, Technical Report SS-98-02*. 1998, pp. 110-114.

22. N. Oliver, E. Horvitz, and A. Garg. "Layered Representations for Recognizing Office Activity." ICMI 2002, pp. 3-.8

23. D. Patterson, L. Liao, D. Fox, H. Kautz. "Inferring High-Level Behavior from Low-Level Sensors". Ubicomp 2003 (to appear).

24. M.E. Pollack, C.E. McCarthy, S. Ramakrishnan, I. Tsamardinos, L. Brown, S. Carrion, D. Colbry, C.Orosz, and B. Peintner. Autominder: A Planning, Monitoring and Reminding Assistive Agent. In 7^{th} *International Conference on Intelligent Autonomous Systems*, March 2002.

25. A. Schmidt, H-W. Gellersen, and C. Merz. "Enabling Implicit Human Computer Interaction - A Wearable RFID-Tag Reader". ISWC 2000,. Pp 193-194.

26. David Tennenhouse, "Embedding the Internet: proactive computing," Communications of the ACM, vol. 43, no. 5, pp. 43-50, May 2000.

27. Q. Tran, K. Truong, and E. Mynatt, "Cook's Collage: Recovering from Interruptions." Demo at Ubicomp 2001.

28. D. Wan, "*Magic Medicine Cabinet: A Situated Portal for Consumer Healthcare*". HUC '99. pp 352-355.

29. WordNet. www.cogsci.princeton.edu/~wn

The columns on the last page should be of equal length.