Interactive Techniques for Labeling Activities Of Daily Living to Assist Machine Learning

Edison Thomaz, Thomas Plötz, Irfan Essa, Gregory D. Abowd
Georgia Institute of Technology, Atlanta, GA

Abstract

Over the next decade, as healthcare continues its march away from the hospital and towards the home, logging and making sense of activities of daily living will play a key role in health modeling and life-long home care. Machine learning research has explored ways to automate the detection and quantification of these activities in sensor-rich environments. While we continue to make progress in developing practical and cost-effective activity sensing techniques, one large hurdle remains, obtaining labeled activity data to train activity recognition systems. In this paper, we discuss the process of gathering ground truth data with human participation for health modeling applications. In particular, we propose a criterion and design space containing five dimensions that we have identified as central to the characterization and evaluation of interactive labeling methods.

Introduction

As we move towards a healthcare system centered on disease prevention, keeping track of Activities of Daily Living (ADLs) in a home environment (e.g. cooking, bathing) becomes increasingly important. In this context, one of the significant challenges involves developing practical, non-invasive, and cost-effective approaches to continuous, personal health monitoring in the home. This is particularly true for older adults in life-long home care.

Recognizing ADLs in a home setting has been an area of intense research work over the last decade. Several approaches that sense humans in the physical world have been demonstrated. Some of these rely on mobile and accelerometer-based sensors [11,12] while others are based on environmental sensing [7,8,9,10]. In the latter case, an environment such as a living room is outfitted with cameras, microphones and direct sensors and a combination of computational techniques is used to predict activity. Environmental and mobile sensing approaches do work well, but typically come with cost, maintenance and privacy implications that render the technique impractical for real-world deployments.

A promising new method for observing human activity in a home environment is called Infrastructure-Mediated Sensing (IMS) [2,3]. The core concept behind IMS is to choose a specific form of home infrastructure (i.e. power lines [3,4], water lines [1,2], gas lines [5], HVAC [6]) and create a device that attaches and monitors the chosen infrastructure with a high degree of fidelity. By sensing some physical phenomenon (e.g., voltage or pressure) that is linked to the chosen infrastructure, it is possible to try to infer what actions are being performed on that infrastructure. For instance, Hydrosense, an IMS-based system around water pressure sensing, has shown that it is possible to link water usage with water fixtures in a single-family home. In other words, by analyzing unique pressure changes in a home's water system, it is possible to estimate which water fixture is in use with a high degree of accuracy.

Despite significant advances in home activity recognition systems, one problem has remained largely unaddressed. How can ADLs be identified if the system has not been taught what a particular activity looks like? To train and test an activity recognition system, it is critical that a set of labeled data corresponding to *ground truth* (in our case the specific human activities we are trying to predict) is available. This is so that an association can be established between a high-level activity and features extracted from the low-level signal.

One approach for obtaining ground truth data is to classify activities without explicit labels. Unsupervised machine learning techniques can be used to cluster activities. In general, however, there is no guarantee that the resulting clusters will correspond to the important activities in a complex environment like the home. The clustering process also does not account for mapping resulting clusters to meaningful and human-interpretable activities. An alternative is semi-supervised learning, in which a portion of the data is labeled and the rest is left unlabeled. This still requires a corpus of labeled data, which has traditionally been collected by instructing people to label their activities as each activity takes place. This is difficult in practice, due to varying scales of activity, inaccuracies caused by human error, and rapid labeling fatigue.

A Criterion for Interactive Labeling Techniques

Although human participation in ground truth data gathering can be a challenge, we believe that bringing individuals into the process is not only immensely more practical than other approaches (e.g. installation of validation sensors), but also offers an opportunity for individuals to engage reflectively about their ADLs. In order to consistently examine, compare and evaluate different labeling approaches with human participation, we have devised a criterion and design space made up of five dimensions: (1) Completeness & Correctness, (2) Human Effort & Motivation, (3) Level of Participation, (4) Frequency, and (5) Temporal Alignment.

Completeness & Correctness

The most obvious approach for obtaining activities of daily living labeled data with human participation is to provide individuals with a logbook or mobile device and ask them to register their activities immediately before, during or after performance of said activities. In theory, complete ground truth is achieved with this technique since all events get labeled. However, there are many reasons why that is not necessarily the case in practice. Even when absolutely all ADLs are labeled, which is an unlikely event, some of the labels might be false positives, or be subjected to substitution error.

Now consider the scenario where individuals provide a high-level account of their activities of daily living only once or twice a day using the Day Reconstruction Method (DRM) survey tool [13]. This approach eliminates the tedious and repetitive task of continuous activity labeling throughout the day, deriving labels from concise activity descriptions of one's day instead. A description might be "*We woke up at 8AM, took a shower, had breakfast and went to the park. We played soccer, had a picnic with friends and came home about 4PM. After a short nap, we started cooking dinner*". From the description, it is possible to have a general idea of when certain activities took place, such as breakfast, and associate them with the underlying physical phenomenon being observed, such as water pressure changes. However it is unlikely that a full day's worth of activities can be reconstructed from such a short description, resulting in activity miscounting and possibly mislabeling too.

Completeness and correctness refer to the likelihood of reaching ground truth in a labeling task when taking into account activity miscounting and mislabeling. The examples above show how difficult it is to achieve absolute completeness and correctness whether individuals label activities in real-time or not. As one would expect, the closer to completeness and correctness, the better the labeling technique.

Human Effort & Motivation

How much effort does an individual have to put into the labeling process? Do individuals have to assign labels as they are performing an activity, or just once a day? Are there any sub-systems that can assist, guide or steer individuals towards a particular label? Is there an element of entertainment associated with labeling, to keep individuals engaged and motivated?

It has been observed in recent years that incorporating gaming elements into the workflow of time-consuming or tedious tasks can be an effective, persuasive technique to keep individuals engaged for a longer period of time. Imagine a Foursquare-style data labeling game where individuals "check-in" on their mobile devices as they perform activities of daily living at home. A scoreboard could be setup to keep track of who performs certain tasks the most or the least, and "awards" could be given to winners. These awards could be exclusive to the virtual environment of the game, or be redeemed for goods and discounts in the physical world. A scenario we don't believe is far fetched is one where a percentage of health insurance cost is tied to performance in a health-at-home game, where activities of daily living serve as a proxy for health and lifestyle inference.

There can be a high cost associated with labeling ADL events. In order to obtain help from individuals in this potentially arduous task, labeling methods might want to consider how to reward and motivate individuals for their work, whether it benefits them directly or not. The amount of effort involved in the labeling process, or return on the labeling investment, is one of the most important parameters and determines whether a particular labeling method is practical and sustainable in the long term.

Level of Participation

In order to label activities, an individual usually needs to be an active participant. But that is not always the case; at times, being an observer of the ADL is enough to properly label it. In some cases, given enough information and context, it might even be possible for activities to be labeled through a crowd-sourced approach, much like the human computation community promotes.

Level of participation is directly related to how much contextual information an individual needs to know in order to help with labeling. A pertinent question here is whether individuals have to recollect the past in order to provide accurate labels. For instance, the longer the time difference between activity and labeling, the more individuals have to know about the specific situation.

In general, the less dependent on a single individual the labeling method is, the more flexible we perceive it to be. An approach that we believe is quite promising is one where a set of individuals who are directly or indirectly involved in an ADL independently label it. The labels are then compared and the most popular one is assigned as ground truth.

Frequency

Consider a labeling technique motivated by active learning, a supervised machine learning approach. Here, the sensing infrastructure queries the individual for labels as new, previously unseen activity events, or clusters of similar events are identified. After an initial period of constant label querying, the number of unseen events falls sharply, reducing queries to a minimum.

Frequency refers to how often individuals are required to label activities. Some techniques demand labeling of ADLs as they take place, while others may require labeling only once or twice a day. Labeling frequency is directly related to the effort involved in labeling, since a high label count requirement over time results in a high labeling workload for participants. Frequency is also inversely proportional to the likelihood of correct and complete labeling. The higher the labeling frequency, the higher the chance that individuals will get fatigued or bored over time and either mislabel or skip labeling altogether.

Temporal Alignment

The reason why labeled ADL data is so valuable in modeling is because it associates a concept, activity or object (e.g. water fixture in the kitchen, taking a shower) with a pattern in a physical modality (e.g. voltage or pressure change) observed at some point in time. Temporal alignment refers to the level of difficulty in temporally synchronizing a label given for a pattern to the pattern itself.

Consider a previously described labeling technique where individuals register activities immediately before, during or after performance of said activities on a logbook or mobile device. Despite the shortcomings of this approach, all labels end up in perfect alignment with the activities, as expressed by some observable physical phenomenon. On the other hand, any labeling technique that relies on the recollection of activities, such as the one employing the Day Reconstruction Method (DRM) survey tool, suffers from temporal alignment since it becomes very difficult to associate labels with activities long after the activities have taken place.

Contribution

Personal health monitoring at home is one of the cornerstones of the future of healthcare. Although we have made substantial progress in the development of systems that can observe and log traces of activities of daily living in a home setting, building predictive models of high-level human activities and health continue to be a challenge. This is due in large part to the difficulty in acquiring labeled data used to train these computational models.

In this paper, we propose a criterion based on five dimensions that can be used to characterize and evaluate interactive labeling methods, where interactive is synonymous with human participation. We have found these dimensions to be particularly applicable in the context of ADLs recognition systems. The dimensions are (1) Completeness & Correctness, (2) Human Effort & Motivation, (3) Level of Participation, (4) Frequency and (5) Temporal Alignment.

All labeling techniques have advantages and disadvantages. While some techniques might prove unsuitable given a set of conditions, they might as well be the best choice under other circumstances. We have found the proposed criterion to be a useful metric in facilitating the process of choosing and designing labeling methods with human participation.

References

1. Froehlich, J., Larson, E., Campbell, T., Haggerty, C., Fogarty, J., and Patel, S. HydroSense: Infrastructure-Mediated Single-Point Sensing of Whole-Home Water Activity. Proceedings of the International Conference on Ubiquitous Computing (UbiComp 2009).
2. Froehlich, J., Larson, E., Saba, E., Campbell, T., Atlas, L., Fogarty, J., and Patel, S. A Longitudinal Study of Pressure Sensing to Infer Real-World Water Usage Events in the Home. Proceedings of the International Conference on Pervasive Computing (Pervasive 2011).
3. Gupta, S., Reynolds, M., and Patel, S. ElectriSense: Single-Point Sensing Using EMI for Electrical Event Detection and Classification in the Home. Proceedings of the International Conference on Ubiquitous Computing (UbiComp 2010).
4. Patel, S.N., Robertson, T., Kientz, J.A., Reynolds, M.S., and Abowd, G.D. (2007). At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line. Proceedings of the International Conference on Ubiquitous Computing (UbiComp 2007).
5. Cohn, G., Gupta, S., Froehlich, J., Larson, E., and Patel, S. GasSense: Appliance-Level, Single-Point Sensing of Gas Activity in the Home. Proceedings of the International Conference on Pervasive Computing (Pervasive 2010).
6. Patel, S., Reynolds, M., and Abowd, G. Detecting Human Movement by Differential Air Pressure Sensing in HVAC System Ductwork: An Exploration in Infrastructure Mediated Sensing. Proceedings of the International Conference on Pervasive Computing (Pervasive 2008).
7. Munguia Tapia, E., Intille, S.S., and Larson, K. Activity Recognition in the Home Using Simple and Ubiquitous Sensors. Proceedings of the International Conference on Pervasive Computing (Pervasive 2004).
8. Intille, S., Munguia Tapia, E., Rondoni, J., Beaudin, J., Kukla, C., Agarwal, S., Bao, L., and Larson, K. Tools for Studying Behavior and Technology in Natural Settings. Proceedings of the International Conference on Ubiquitous Computing (UbiComp 2003).
9. Fogarty, J., Au, C., and Hudson, S.E. Sensing from the Basement: A Feasibility Study of Unobtrusive and Low-Cost Home Activity Recognition. Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2006).
10. Wilson, D., Consolvo, S., Fishkin, K., and Philipose, M. In-Home Assessment of the Activities of Daily Living of the Elderly. Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2005): Workshops-HCI Challenges in Health Assessment.
11. Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemingway, B., Hightower, J., Klasnja, P., Koscher, K., LaMarca, A., Landay, J., and Lester, J. The Mobile Sensing Platform: An Embedded System for Capturing and Recognizing Human Activities. Proceedings of the International Conference on Pervasive Computing (Pervasive 2008).
12. Consolvo, S., McDonald, D., Toscos, T., Chen, M., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., and Landay, J. Activity Sensing in the Wild: A Field Trial of UbiFit Garden. Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2008).
13. Kahneman, D., Krueger, A., Schkade, D., Schwarz, N., and Stone, A. A Survey Method For Characterizing Daily Life Experience: The Day Reconstruction Method (DRM). Mimeo, Princeton University, 2003.