# Detecting Health and Behavior Change by Analyzing Smart Home Sensor Data

Gina Sprint, *Student Member, IEEE*, Diane J. Cook, *Fellow, IEEE*, Roschelle Fritz, and Maureen Schmitter-Edgecombe

***Abstract*— Smart home environments offer an unprecedented opportunity to unobtrusively monitor human behavior. Sensor data collected from smart homes can be labeled using activity recognition to provide behavior context and to help determine whether relationships exist between behavior in the home and health changes. To detect and analyze behavior changes that accompany health events, we introduce the behavior change detection (BCD) approach. BCD detects activity timing and duration changes between windows of time, determines the significance of the detected changes, and analyzes the nature of the changes. We demonstrate our approach using three case studies for older adults living in smart homes who experienced major health events, including cancer treatment, insomnia, and a fall. Our results detect behavior changes consistent with the medical literature for these cases. The results suggest the changes can be automatically detected using BCD. The proposed smart home, activity recognition algorithms, and change detection approach are useful data mining techniques for understanding the behavioral effects of major health conditions.**

## I. Introduction

In recent years, sensors have become ubiquitous in our everyday lives. Sensors are ambient in the environment, embedded in smartphones, and worn on the body. Data collected from sensors form a time series in which each sample of data is paired with an associated timestamp. This sensor-based time series data is valuable when detecting and analyzing changes associated with human behavior changes, such as seasonal variations, new lifestyle choices, or new job situations. Sensor-based time series data can also be analyzed to monitor changes related to health events such as a fall, cancer treatment, or onset of a chronic medical condition. Automatically tracking behavior changes from sensor data can help with understanding the behavior impact of these health events. Similarly, detecting these changes can alert individuals and their caregivers about potential health concerns.

In this paper, we introduce a method to analyze the behavioral impact of health events using smart home sensor data called Behavior Change Detection (BCD). Smart home sensor systems provide the capability to automatically and unobtrusively collect information about a resident's everyday behavior. We collect data from ambient sensors placed in smart home environments and label the data with the corresponding activities using automated activity recognition. To track changes in routine behavior, we quantitatively compare two or more time periods, or windows, of activity-labeled data. If the two time windows contain significantly different activity information then this may indicate a significant behavior change, possibly due to a health event [1]. The ability to be physically active is an indicator of good health [2] but the opposite is also true, inactivity or a change in activity can be an indicator of a change in health due to chronic or acute illness [3].

We hypothesize that the relationship between sensed behavior and health events can be observed and analyzed using smart home data. To evaluate our hypothesis, we utilize BCD to analyze smart home data collected for multiple years in the homes of older adults. Health events are identified for three of the smart home residents based on medical records review and monthly interviews with the study participants. Data surrounding the health event is compared with baseline normal data to determine if a significant behavior change has occurred and describe the nature of the change. The corresponding behavior change is then analyzed by a clinician to validate the behavior change and explain the relationship between the health event and corresponding behavior change. Results from the case studies presented in this paper indicate that smart home and machine learning technologies can be used to understand the behavioral impacts of health events and to provide information indicating possible health concerns.

## II. Related Work

Similar research work in this area consists of activity recognition approaches and research in activity change detection. Activity recognition algorithms have been designed for wearable, phone, home, video, and other sensors using machine learning techniques that range from naïve Bayes classifiers and decision trees to more complex models including Gaussian mixture models and conditional random fields [4], [5]. Most of the prior work in detecting behavior changes utilized wearable data to correlate home-based movement with health measures [6], [7], although smart home data have been used to analyze mobility and time out of the home with respect to cognitive and physical health [8], [9]. Our own earlier work showed that smart home data can be analyzed over time to predict performance on cognitive health assessment tests [10].

This work is funded in part by National Science Foundation grant 0900781 and National Institutes of Health grant R01EB015853.

Gina Sprint and Diane J. Cook are with the School of Electrical Engineering and Computer Science at Washington State University, Pullman, WA 99164 USA (e-mails: gsprint@eecs.wsu.edu, cook@eecs.wsu.edu).

Roschelle Fritz is with the College of Nursing at Washington State University-Vancouver, Vancouver, WA 98686 USA (e-mail: shelly.fritz@wsu.edu).

Maureen Schmitter-Edgecombe is with the Department of Psychology at Washington State University, Pullman, WA 99164 USA (e-mail: schmitter-e@wsu.edu).

## III. Methods

### A. Smart Home Sensor Data

We collect data in everyday home environments using the CASAS "smart home in a box" [11]. The three homes that we include in this study are single-resident apartments. These homes are equipped with combination motion/light sensors on the ceilings and door/temperature sensors on cabinets and doors. The sensors continuously and unobtrusively monitor daily activities of the residents by sending text message-type updates, or sensor events, whenever they sense a state change (i.e., from "door closed" to "door open" or from "no motion" to "motion"). Table I shows example sensor data.

Once the sensor data is collected, each sensor event is labeled with the corresponding activity using the CASAS-AR activity recognition algorithm [12]. Let $A = \{a_1, a_2, \ldots, a_T\}$ be the set of all activities. Given features $x_t \in \mathbb{R}^d$ extracted from a sequence of sensor events ending at time t, the challenge of activity recognition is to map $x_t$ onto a value $a \in A$ indicating the activity that occurred at time $t$. These labels provide a vocabulary for expressing and analyzing the sensed behavioral patterns. In this study we analyze the activities of Hygiene, Sleep, Bed-Toilet, Eat/Drink, Enter/Leave Home, Relax, and Work. CASAS-AR labels sensor events with activity labels in real time as the events occur. To do this, a dynamic-size sliding window is moved over the sensor events and features $x_t$ describing the current window of information are extracted. The computed activity recognition features include the sensor event time of day, the size of the sliding window, the event count for each sensor within the window, time elapsed for each sensor since its most recent event, the most recent event location and sensor identifier, and the sensor generating the most events in the previous two windows.

Training data for CASAS-AR are provided by external annotators who look at one month of data and utilize both the house floorplan and resident information to generate corresponding ground truth activity labels. Using this method, CASAS-AR learns an activity model based on training data from multiple smart home sites and can thus generalize for application to new smart homes with no training data. For these activities in the three smart home testbeds we analyze in this paper, CASAS-AR achieved a recognition accuracy of 98% using 3-fold cross validation.

### B. Behavior Change Detection

We are interested in analyzing the behavioral impact of health events. More specifically, we want to determine if a significant change in behavior has occurred at the time of the health event and to analyze the nature of the behavior change. To do this, we introduce methods to quantify the amount of change in activity patterns between two windows of time series activity data that were sampled by smart home sensors and labeled by CASAS-AR. Let $X = \{d_1, d_2, \ldots\}$ denote a sample of time series data where each day's data are expressed by a vector of extracted activity features $d$. Let $W$ be a window of $n$ days such that $W \subseteq X$. Daily activity features are computed for each day in $W$, including:

- Amount of time spent on each activity.

TABLE I

EXAMPLE SENSOR DATA AND ACTIVITY LABELS

| Timestamp/Identifier/Message | Sensor Location | Activity |
|---|---|---|
| 2014-06-15 03:38:28.094897 M009 ON | BedroomMotion | Sleep |
| 2014-06-15 03:38:29.213955 M009 OFF | BedroomMotion | Sleep |
| 2014-06-15 03:38:17.814393 M015 ON | BathroomMotion | Bed-Toilet |
| 2014-06-15 03:38:58.584179 M015 OFF | BathroomMotion | Bed-Toilet |
| 2014-06-15 03:39:17.814393 M009 ON | BedroomMotion | Sleep |

- Sensor density of each activity (measured as number of sensor events).
- Total amount of movement that occurs in the home, expressed as the total distance travelled.

BCD compares two windows of data, $W_i$ and $W_j$, within time series $X$. We set the window size to one week in length ($n = 7$). BCD compares a baseline window ($i = 1$, the first week in our dataset representing normal behavior for the resident) with each subsequent window ($j = 2,3, \ldots$). We utilize two change detection methods, sw-PCAR and virtual classifier, to offer two slightly different perspectives on the data comparison.

#### 1) sw-PCAR

Our Permutation-based Change Detection in Activity Routine (PCAR) approach [10] was originally designed to analyze changes in longitudinal smart home data. Here we adapt the original approach to handle smaller windows of activity-labeled data. The resulting small-window PCAR (sw-PCAR) algorithm segments each day within the window into non-overlapping hour-long time intervals. Each time interval has a corresponding probability distribution over the activities that occur at that time. Days within two windows $W_i$ and $W_j$ are averaged to yield aggregate windows $\widehat{W_i}$ and $\widehat{W_j}$. Next, we compute a change score $CS$ using the symmetric Kullback-Leibler (KL) divergence distance between the activity probability distributions in $\widehat{W_i}$ and $\widehat{W_j}$.

The significance of the distance value $CS$ is determined by concatenating windows $\widehat{W_i}$ and $\widehat{W_j}$ into one window $W$. All of the time intervals within $W$ are randomly shuffled then split into two new sub-windows and the KL distance is computed for this permuted window pair. This shuffling procedure is repeated N times to produce a N-length vector $V$ of KL distances. If N is large enough, the corresponding set of KL distances forms an empirical distribution of the possible permutations of activity data for the two windows. $CS$ is then compared to the permutation vector $V$ using boxplot-based outlier detection. To determine if $CS$ is an outlier of $V$, the interquartile range (75$^{th}$ percentile – 25$^{th}$ percentile) of $V$ is computed. If the value of $CS$ is outside of the $1.5 \times 75^{th}$ percentile, $CS$ is considered an outlier and thus a significant change score.

#### 2) Virtual Classifier

The second method, virtual classifier (VC), utilizes a binary classifier to detect and explain behavior change. This type of virtual classifier for change analysis was first proposed by Hido *et al*. [13]. For the VC approach, feature vectors from window $W_i$ are labeled with a positive class and feature vectors from window $W_j$ are labeled with a negative class. VC trains a decision tree to learn a boundary between the virtual positive and negative classes. The resulting average prediction accuracy based on k-fold cross

TABLE II
CHANGE SCORE RESULTS

| Participant | Method | $W_1/W_2$ (baseline) | $W_1/W_{event}$ (health event) |
|---|---|---|---|
| SH1 | sw-PCAR | 0.001 | 0.091* |
| | VC | 0.500 | 1.000* |
| SH2 | sw-PCAR | 0.004 | 0.042* |
| | VC | 0.438 | 1.000* |
| SH3 | sw-PCAR | 0.000 | 0.001 |
| | VC | 0.500 | 0.750* |

SH = smart home, VC = virtual classifier, * = significant result.

validation is represented as $p_{VC}$. If a significant change exists between $W_i$ and $W_j$ then the average classification accuracy $p_{VC}$ of the learner should be higher than the accuracy expected from random noise which is $p_{rand} = 0.5$. The inverse survival function of a binomial distribution can be used to determine a critical value, $p_{critical}$, at which $n$ Bernoulli trials are expected to exceed $p_{rand}$ at $\alpha = 0.05$ significance. If $p_{VC} > p_{critical}$ then a significant change exists between windows $W_i$ and $W_j$.

If the sw-PCAR or VC tests conclude that a change score is significant, the next step is to investigate the data to explain the source of change. Typically this requires computing features that provide a meaningful context for change and applying statistical tests for those specific features. One of the advantages of the VC approach includes its output of a decision tree learner. The learner provides an explanation of the source of change without reliance on statistical tests. Upon detecting a significant change, the decision tree is retrained on the entire dataset and inspected to reveal the features that are the most valuable in discriminating between the two windows of data.

### C. Smart Home Resident Health Events

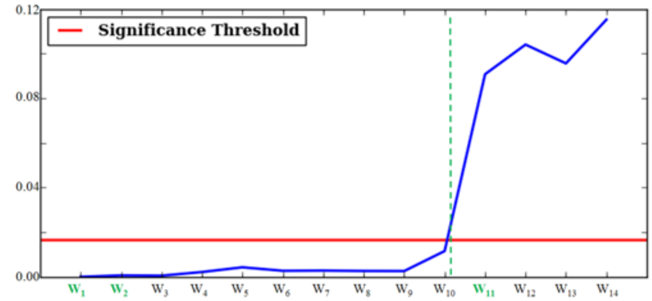
We collected data in smart homes with older adult residents for multiple years. For each smart home resident we also recorded health events with their date and event type, based on medical records and monthly interviews with the participants. Here we describe three of these health events and utilize these case studies to illustrate the use of BCD:

- SH1 (86 year old female): Diagnosed with lung cancer and started radiation treatment during week $W_{10}$.
- SH2 (91 year old female): Diagnosed with insomnia during $W_{11}$.
- SH3 (80 year old female): Fell in her home during $W_8$.

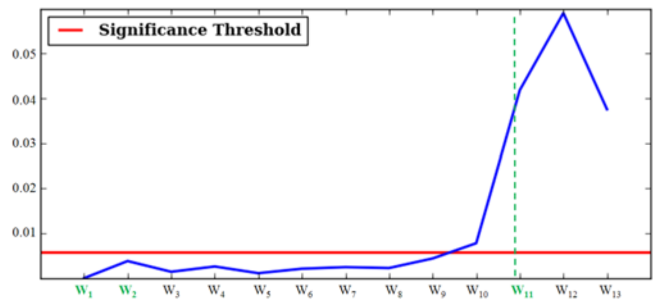
We hypothesize that each of these life-changing health events will have an observable and quantifiable impact on the corresponding participant's behavior. To validate this hypothesis, we use BCD to compare one-week baseline of smart home activity data ($W_1$) with two other weeks. The first comparison is with another pre-event week, namely the week immediately following the baseline ($W_2$). The second comparison is with the first full week during which the individual experienced the health event ($W_{event}$).

## IV. RESULTS

The change scores using the two BCD techniques described in this paper (sw-PCAR and virtual classifier) are summarized in Table II. Fig. 1 illustrates weekly results from applying sw-PCAR change detection for SH1 and SH2. Fig. 2 shows associated activity density maps for select activities for each participant. Density maps have been used in prior work to visualize levels of movement in the home [9]. Our activity density map is a heat map that visualizes the amount of time spent on a particular activity as a function of a 24-hour clock (y axis), aggregated over one week (x axis). The darker the color, the more time was spent on the activity during that particular hour of the day in the corresponding week.



a) SH1



b) SH2

Fig. 1. Smart home resident SH1 and SH2 sw-PCAR change scores for baseline week $W_1$ and health event week $W_{11}$. Values above the red line are significant changes. The green dashed line indicates the occurrence of the health event.

## V. DISCUSSION

In this paper, we utilize three smart home resident case studies to illustrate the use of BCD for detecting major health events. For participant SH1, the behavior changes during radiation treatment are evident for each of the change detection methods and the results are significant (see Table II and Fig. 1a). As the density maps show in Fig. 2a and 2b, the participant's level of sleep decreased once treatment started and the number of times she left the home / returned home increased. Furthermore, the top-level decision tree feature for SH1 is the number of sensor events that are related to an "Enter Home" activity. Possible explanations for this are increased trips out of the home for treatment appointments or visits from family and caregivers. Another impact of the treatment is the increased number of trips this participant made to the kitchen to eat or drink. These more frequent kitchen trips are consistent with the observation that radiation treatment increases the feeling of thirst, resulting in a patient drinking more liquids throughout the day [14].

SH2 is a 91 year old female smart home resident diagnosed with insomnia. The change scores for SH2 reported in Table II indicate significant changes in overall routine are detected by both sw-PCAR and VC methods. In

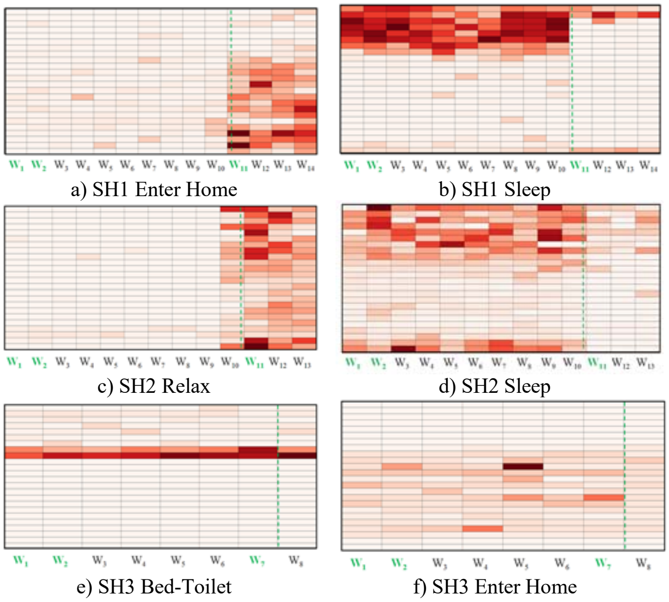



Fig. 2. Activity density maps for smart home residents SH1-3. The green dashed line indicates the occurrence of the health event.

Fig. 1b we see that changes occur not only during week $W_{11}$ but also in the days leading up to the health event and persisting to days and weeks following the insomnia diagnosis. We can also observe in the density maps (Fig. 2c and 2d) that the amount of sleep does decrease during this period. The change in behavior also impacts relaxation, which is time spent in a favorite chair or couch with little movement and possibly napping. In addition, the number of trips outside the home decreases during this time. The virtual classifier actually finds the corresponding decrease in "Enter Home" events to be the main discriminating feature between baseline and health event weeks.

The last case study, SH3, experienced a fall during the time that we were collecting sensor data in their home. She described that her right leg hurt for several days after the fall and consequently it "slowed her down." As the results in Table II indicate, this health event has a much more subtle impact on behaviors, at least those that can be detected by ambient smart home sensors. sw-PCAR detects almost no change between weeks $W_1$ and $W_2$ or between weeks $W_1$ and $W_8$. The virtual classifier does detect the change during the health event week. The VC-generated rule indicates the difference is primarily detected based on the total distance that the individual traveled throughout the home on a daily basis. Furthermore, as the density map in Fig. 2f indicates, there is an apparent slight decrease in trips in/out of the home. The decrease in movement is consistent with the observation that the hurt leg caused the resident to slow down.

## VI. Conclusion

In this paper we introduce BCD, an approach to behavior change detection. We describe how BCD can be used to quantify and explain changes that are detected in daily activity data. In particular, BCD can detect changes in smart home-detected behavior data that occur as a result of health events. From the three case studies that we analyzed in this paper we see that the ability to detect behavioral impact of health events depends on the nature of the health event itself. Some events impact multiple activities including sleep, eating, and trips out of the home. In contrast, other events have more localized impact. The ability to detect the actual health event occurrence (e.g., fall) may require more sensitive sensors to be placed in the home or on the body. These insights may help care providers respond to the needs of individuals who are experiencing changes in their health. An algorithm such as BCD can periodically look for changes in behavioral routine and alert the individual and their caregiver about these changes as they may indicate changes in cognitive or physical health. Because BCD can analyze any type of sensor data, our continued research will adapt these methods to analyze smart phone and wearable data, as well as data collected in smart homes.

## References

[1] I.-M. Lee, E. J. Shiroma, F. Lobelo, P. Puska, S. N. Blair, P. T. Katzmarzyk, and Lancet Physical Activity Series Working Group, "Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy," *Lancet Lond. Engl.*, vol. 380, no. 9838, pp. 219–229, Jul. 2012.

[2] "Physical Activity | Healthy People 2020." [Online]. Available: https://www.healthypeople.gov/2020/topics-objectives/topic/physical-activity.

[3] J. T. Cavanaugh, T. D. Ellis, G. M. Earhart, M. P. Ford, K. B. Foreman, and L. E. Dibble, "Toward Understanding Ambulatory Activity Decline in Parkinson Disease," *Phys. Ther.*, vol. 95, no. 8, pp. 1142–1150, Aug. 2015.

[4] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, pp. 1–33, Feb. 2014.

[5] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-Based Activity Recognition," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 6, pp. 790–808, Nov. 2012.

[6] G. LeBellego, N. Noury, G. Virone, M. Mousseau, and J. Demongeot, "A model for the measurement of patient activity in a hospital suite," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 92–99, Jan. 2006.

[7] H. H. Dodge, N. C. Mattek, D. Austin, T. L. Hayes, and J. A. Kaye, "In-home walking speeds and variability trajectories associated with mild cognitive impairment," *Neurology*, vol. 78, no. 24, pp. 1946–1952, Jun. 2012.

[8] S. Robben, M. Pol, and B. Kröse, "Longitudinal ambient sensor monitoring for functional health assessments: a case study," 2014, pp. 1209–1216.

[9] S. Wang, M. Skubic, and Y. Zhu, "Activity Density Map Visualization and Dissimilarity Comparison for Eldercare Monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 607–614, Jul. 2012.

[10] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Modeling patterns of activities using activity curves," *Pervasive Mob. Comput.*, 2015.

[11] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: A Smart Home in a Box," *Computer*, vol. 46, no. 7, pp. 62–69, Jul. 2013.

[12] N. C. Krishnan and D. J. Cook, "Activity Recognition on Streaming Sensor Data," *Pervasive Mob Comput*, vol. 10, pp. 138–154, Feb. 2014.

[13] S. Hido, T. Idé, H. Kashima, H. Kubo, and H. Matsuzawa, "Unsupervised Change Analysis Using Supervised Learning," in *Advances in Knowledge Discovery and Data Mining*, T. Washio, E. Suzuki, K. M. Ting, and A. Inokuchi, Eds. Springer Berlin Heidelberg, 2008, pp. 148–159.

[14] P. Beach, B. Siebeneck, N. F. Buderer, and T. Ferner, "Relationship between fatigue and nutritional status in patients receiving radiation therapy to treat lung cancer," *Oncol. Nurs. Forum*, vol. 28, no. 6, pp. 1027–1031, Jul. 2001.