Context-Aware Delivery of Ecological Momentary Assessment

Samaneh Aminikhanghahi, Maureen Schmitter-Edgecombe, and Diane J. Cook, *Member, IEEE*

***Abstract*— Ecological Momentary Assessment (EMA) is an in-the-moment data collection method which avoids retrospective biases and maximizes ecological validity. A challenge in designing EMA systems is finding a time to ask EMA questions that increases participant engagement and improves the quality of data collection. In this work, we introduce SEP-EMA, a machine learning-based method for providing transition-based context-aware EMA prompt timings. We compare our proposed technique with traditional time-based prompting with for 19 individuals living in smart homes. Results reveal that SEP-EMA increased participant response rate by 7.19% compared to time-based prompting. Our findings suggest that prompting during activity transitions makes the EMA process more usable and effective by increasing EMA response rates and mitigating loss of data due to low response rates.**

***Index Terms*— Activity recognition, Activity transition, Change detection algorithms, Ecological Momentary Assessment (EMA).**

I. INTRODUCTION

RETROSPECTIVE self-reports are the most common assessment method found in clinical psychology [1]. Traditionally, self-report information is gathered through questionnaires which are limited because the details of previous experiences cannot be consistently recalled via retrospective memory [2]. Ecological Momentary Assessment (EMA), or experience sampling, addresses these issues by collecting information about daily experiences at the moment they occur. Advances in smart devices and pervasive computing has made EMA more practical and commonplace in recent years. App-based EMA systems have been designed that allow participants to enter data directly and record the time delay between the EMA question and participant response [2].

Although app-based EMA enables data collection to be simpler and more ecologically valid, participants need to repeatedly respond to prompts and questions which may reduce the overall response rate and number of individuals willing to participate in the study [2][3]. Typical response rates vary depending on the EMA questions and participant groups. Examples are 55% for patients preparing for bariatric surgery [4], 75% when assessing a novel evaluation tool for curricular change in an internal medicine residency program [5], and 78% for effects of Alcohol-Tobacco Co-Use [6].

To overcome an EMA response rate barrier, researchers try to encourage participants to achieve a specified threshold value (75% [7] and 80% [8]) with monetary rewards as an incentive. Yet, some participants do not reach the threshold level and researchers must eliminate part of the collected data in their analysis to reach the predefined response rate. Finding the best time to send prompts and reduce interruption of an individual's normal activity routine is a primary challenge in designing EMA technologies. Receiving notifications at inopportune times is not only annoying, but such interruptions can increase a person's cognitive load [9], distract people from what they are doing, and reduce willingness to interact with the technology [10][11]. Additionally, these difficulties may make participants more hesitant to initiate EMA responses and reduce acceptance of the EMA by decreasing response rates.

To increase the EMA response rate and avoid overwhelming individuals with notifications at unwanted moments, an EMA notification system needs to be aware of the person's behavioral context and detect opportune moments to deliver information and questions [12]. A suitable time to deliver prompts is the time in which they cause minimal interruptions to the user's engaged task [9]. We hypothesize that sending notifications and prompts during activity transitions, or the period of time between the end of one activity and the beginning of the next, can improve response rates, reduce cognitive load, and increase user acceptance of EMA technology. Furthermore, we hypothesize that these activity transitions can be identified independently of the specific activities that are performed, using unsupervised change point detection algorithms.

In this paper, we propose a context-aware machine learning technology, SEP-EMA, to identify the best time to deliver prompts to individuals and ask them to respond to ecological momentary assessment questions. The main objective of SEP-EMA is increasing the response rate to EMA questions regardless of phenomenon of interest and question type. To accomplish this objective, SEP-EMA detects activity transitions in real time and combines transition detection with other contextual information to select prompt times. We validate our hypotheses as part of an in-home study in participant smart homes. A controlled user study with 19 participants shows providing queries at context-sensitive times via SEP-EMA resulted in a 7.1% increase in response rate, compared with queries presented at random times.

A. Ecological Momentary Assessment (EMA)

Ecological momentary assessment (EMA) is a real-time and real-world data collection method which tries to avoid retrospective biases and maximize ecological validity. In recent years, many studies employ EMA methods to characterize individual differences, describe natural history, assess contextual associations, and document temporal sequences [1].

To collect more accurate and ecologically-valid data, EMA asks participants about their experiences closer to the time or event of interest. EMA data sampling can be 1) event-based or event contingent (e.g. after physical activity), 2) interval-based or interval contingent (such as once-a-day diaries), and 3) fixed

This research has been supported in part by National Science Foundation grant 1543656, National Institutes of Health grant R01EB009675, and DOD Award No. AZ150096.

Samaneh Aminikhanghahi and Diane J. Cook are with the School of EECS, and Maureen Schmitter-Edgecombe is with the Department of Psychology, Washington State University, Pullman, WA 99163 USA (e-mails: s.aminikhanghahi@wsu.edu, djcook@wsu.edu, schmitter-e@wsu.edu).

or random time-based prompting (sometimes also called signal contingent) [13].

When delivering EMA questions, researchers often utilize existing app-based platforms or custom software. While these apps are powerful, most of these, such as iHabit [14] and ULTEMAT [15], do not yet provide contextual guidance to improve EMA response rates. iHabit [14] is an iOS-based platform to support EMA data collection. The app sends alerts randomly during study-specified hours. App users respond by answering a series of ''check-in'' questions. Similarly, ULTEMAT [15] is an Android EMA platform which delivers EMA questions at either pre-scheduled or random times.

Existing app-based platforms require clinicians to define the timing of prompts at the beginning of the study, rather than letting the app adapt to an individual's own routine. Here, we introduce SEP-EMA, a new EMA app which partners with the CASAS smart home [16] developed at Washington State University. The EMA app provides prompts to the user on an iPad device within a smart home. Screen shots of the app are shown in Figure 1. The app communicates with the smart home through the CASAS server infrastructure and can send queries at context-sensitive times as well as random times.

When a prompt is requested, the server sends a command to the EMA iPad app to prompt the user. When this occurs, the iPad will display the first prompt question and play an alert tone to get the user's attention. This tone is repeated over the course of ten minutes while the same prompt is continuously displayed on the iPad. If the user does not interact with the iPad during the ten minutes (or starts interacting, but then does not finish the prompt within the following ten minutes), the prompt times out and we record this as a no-response prompt.

When the user is prompted, the iPad shows three pages of questions and brings up a cognitive exercise to complete (i.e., a 45 second n-back test [17]). These EMA queries, shown in Figure 1, include: 1) Prompt timing question: This prompt asks the user whether the timing of the EMA questions was Convenient, Neutral, or Poor and used to compare the random and EMA-detected prompt conditions, and 2) EMA assessment questions: This app page consists of seven EMA questions to understand the subject's current mental and physical state. For each question, the user is given a choice of five responses from "Not at all" to "Very much".

In addition to these two pages of questions, SEP-EMA asks participants to label the activity they were recently performing (for validation of activity recognition) and brings up a n-back test (i.e., indicating whether a shape was the same or different from the prior presented shape by quickly pressing a "yes" or "no" button) to measure the individual's cognitive functioning. In this paper, we do not focus on the outcome of the EMA assessment, activity validation, or the n-back test. Instead, we focus on the prompt timing question to understand whether our proposed algorithm increases the response rate of a participant by being sensitive to their activity patterns and current context.

B. Related Work

A growing body of work documents use of EMA for assessment as well as a need for techniques that improve design of the timing and delivery for EMA queries. Researchers have developed software applications for collecting ecological momentary data that analyze different psychological behavior such as depressive symptoms [18][19], alcohol use disorder [20][21], physical activity [22][23], and smoking [24][25].

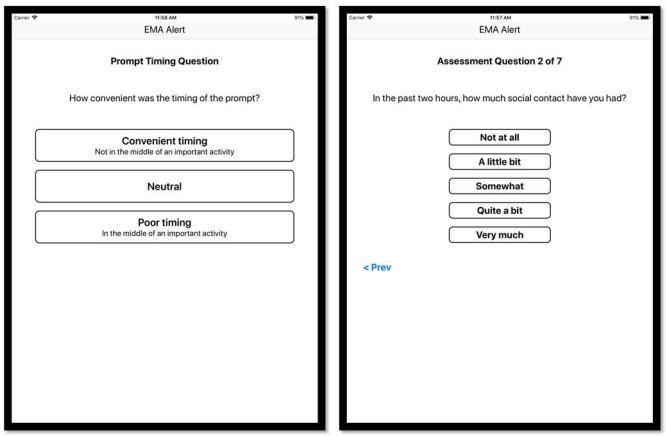

Fig. 1. EMA app prompt timing and assessment question page.

To reduce the burden on participants and encourage them to answer EMA questions more frequently, study designers need to carefully define the times that the EMA queries are posed to individuals. A majority of approaches employ fixed time-triggered scheduling. A time-triggered EMA app delivers prompts based on a pre-specified, inflexible time or a random time within a predefined time window. For example, Ballegooijen et al. [18] sent one to three notifications through an individual's smartphone at a random time point between ten o'clock in the morning and ten o'clock in the evening asking participants about their depressive symptoms. Similarly, Mundi et al. [4] sent questions to pre-bariatric surgery patients that asked about lifestyle based on key patterns entered by the subject at enrollment: wake-up time; breakfast, lunch, and dinner times (weekday and weekend); and sleep time.

Time-triggered scheduling notifications may be delivered when the individual is engaged in another important task. Receiving prompts, notifications, or questions during such times can increase a person's cognitive load and introduce errors into their current task errors [26]. Furthermore, sending messages on predefined timeframe may preclude participants responding to a signal when it occurs (e.g., if they are taking a shower when the prompt arrives). To overcome these issues and improve the performance of EMA, context-aware triggering can be used. A context-aware EMA delivers prompts and messages to individuals at times that are sensitive to the person's status, activity, and environment.

Because of increased awareness of the need for context sensitivity, researchers are now starting to consider context-aware prompting for smartphone information notification management. Smartphone technology has become ubiquitous. The escalated dependence on these devices causes an increase in interruptions due to phone calls, texts, reminders, prompts, and notifications. In response, app designers attempt to defer notifications to times that maximize the probability of responding. To identify the proper time, study designers manually create complex timing rules based on current activity, the transition between activities, contextual data, or a combination of these factors [27].

As evidence of the relationship between a person's activity and their ability to process and respond to interruptions, Gillie

et al. [28] found that interruptions are particularly harmful if the user's current activity is similar to the topic of the interruption. This may occur, for example, when both the current task and the interruption involve processing numeric information. The nature of the current activity is also important: interruptions may be worse when the person is taking part in a more distractible activity. Activities that require frequent access to working memory such as paying bills are considered distractible activities [29]. Rosenthal et al. [30] demonstrated that smartphone users also exhibit varying preferences and costs associated with interruptions, even while performing a similar activity.

This amount of change in response to interruptions indicates the need for personalized interruption management. Researchers have effectively applied machine learning techniques to a variety of biomedical informatics challenges [31][32]. Machine learning-based personalization has also been considered as described by Aminikhanghahi et al. [33], but this supervised learning approach requires a substantial amount of labeled training data for each individual.

On the other hand, there is evidence that people are more accepting of notification interruptions when they have finished one task and have not yet started another [34]. Fischer et al. [35] suggested that notifications which are delivered at the endings of episodes of mobile interaction (making voice calls or texting) have a higher chance of eliciting a response. Similarly, Okoshi et al. [36] proposed "Attelia" as a smartphone middleware to detect breakpoints of a user's mobile interactions in real time and defer notifications until such a breakpoint occurs.

Pejovic et al. [10] developed an interruption management library called InterruptMe for Android-based mobile devices. InterruptMe constructs intelligent interruption models based on a series of machine learning algorithms for interruptibility prediction. Here, smartphone sensors collect contextual information and InterruptMe then performs experience sampling to ask users about their interruptibility at different moments and contexts. As an example of another such management tool, Park et al. [37] introduced SCAN, a social context-aware smartphone notification management system which delivers notifications at appropriate breakpoints. SCAN uses beacons (Bluetooth radio transmitters) and a microphone to detect the presence of conversations as well as the current activity. Then the method employs a decision tree to identify opportune breakpoints for delivering all queued notifications.

However, all of these notification management systems incorporate smart phone data and consider only the interaction of individuals and their phones as the context for detecting breakpoints. In this paper, we introduce SEP-EMA, a context-aware EMA delivery system that partners a mobile device-based EMA delivery mechanism with smart home-based awareness of an individual's activities to find natural breaks in a person's entire daily routine. SEP-EMA uses activity breakpoints (transitions between daily activities) as candidate times for prompting. We then finalize the decision by considering the contextual data together with the current time and current activity.

II. METHODs

A. Design

SEP-EMA is an intelligent prompting model that takes advantage of pervasive computing, signal processing, and machine learning techniques to provide personalized, activity context-aware EMA prompt timings in smart homes. Using embedded sensors, smart homes collect information about the state of the home and the resident(s) to monitor and analyze daily activities. Sensors generate "events" to report their state. These events are triggered, or generated, whenever the sensor detects a change in its state. Each sensor event, e, takes the form $e=< t, s, m >$ where t denotes the timestamp when the sensor message was collected, s denotes the identifier of the sensor generating the message, and m denotes the sensor message.

Figure 2 provides an overview of our proposed EMA question delivery model. When a sensor event is triggered, the information is sent as an input to the feature extraction algorithm. The feature extraction algorithm slides a *window* over the sensor data that looks at 30 consecutive events (ending with the most recent event) and extracts a corresponding feature vector. These features include window time duration, most recent event time, the entropy-based complexity of events within the window, the number of transitions between areas of the home, the change in overall activity level between the first and second halves of the window, the most frequently-triggered sensor, the number of events for each sensor in the current window, and the elapsed time since each sensor fired. Thus, the feature vector dimension varies depending on home floorplans and number of sensors placed in the home. The extracted feature vector is then fed to both the activity transition detection and the activity recognition algorithms. The full set of features that are extracted for activity recognition and activity transition detection is shown in Table I.

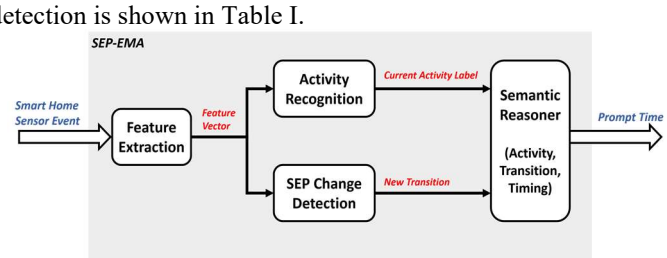

Fig. 2. A system component overview of SEP-EMA.

TABLE I
FEATURE SETS USED IN ACTIVITY TRANSITION DETECTION AND RECOGNITION

Domain	Feature Name
Time features	day of week, hour of day, seconds past midnight
Window features	window duration , most recent sensor, last sensor location, current and previous frequent sensors, entropy-based complexity, activity level change between window halves, number of transitions
Sensor features	count of events for each sensor, elapsed time for each sensor since last event

After completing feature extraction, we apply the change detection algorithm to determine whether there is an activity transition or breakpoint within a sequence of sensor data. Change point detection (CPD) is the problem of identifying each time point t when the probability density function f created from sensor data observed before t is sufficiently different from the data observed immediately after t, based on a change in the type of the function or the parameters characterizing the function. We use an approach introduced in our previous work

for activity transition detection called SEP [38]. SEP is a non-parametric sequential change detection algorithm which calculates a change score between two consecutive windows of data (sequences of sensor events) to indicate the amount of change that occurs from the first window to the next. To distinguish this window from one used for feature extraction, which may be different sizes, we will refer to the window used in SEP as a *change point view*. SEP utilizes separation distance as the CPD dissimilarity measure and estimates the probability distribution ratio between the two views surrounding a candidate change point. The separation distance S can be calculated as shown in (1).

$$S = Max_i(1 - \frac{f_{t-1}^i(x)}{f_t^i(x)}) \quad (1)$$

$f_{t-1}^i(x)$ and $f_t^i(x)$ are the estimated probability densities of the first and second view, with length n. The variable i iterates over all data points in each view. To estimate the probability densities, we use cross-validation over data points within each change point view. The maximum of these estimated probability density ratios is then computed. In order to calculate separation distance as a change point score, we estimate the ratio of the two probability densities by a kernel function $g_i(x)$ and calculate the change point score, $\widehat{SEP}$, as shown in (2) [38]. The result is compared with 0 to avoid negative values.

$$\widehat{SEP} = \text{Max}\left(0, \left(1 - \frac{1}{n}\sum_{i=1}^{n} g_i(x)\right)\right) \quad (2)$$

The score calculated by (2) can be used to detect change points. Since a larger SEP score means that the probability of a change point is greater, we reject all candidate points whose SEP values are lower than a threshold value. To reduce the chance of false alarms and avoid double change points (two change points in quick succession that are part of the same transition), we only consider local peak score values as change points. SEP-based transitions are not detected exactly in real time because the algorithm requires access to two sensor events at future times t_{i+1} and t_{i+2} to decide if there is a transition at time t_i. We therefore refer to this algorithm as operating in *2-real time*. Here, ε=2 denotes the number of sensor events after the candidate change point that must be processed. The output of the algorithm is a tag attached to each sensor event indicating whether the event is a transition or not.

The extracted feature vector is also used for activity recognition. We employ the AR algorithm [39] that utilizes a Random Forest (RF) classifier with 100 decision trees to label sensor events. To establish the performance of AR, we evaluate it using data collected in 30 smart home testbeds from a previous study [16]. We use these testbed smart homes for evaluation of the activity recognition and change point detection components used in SEP-EMA because ground truth activity labels were available for these testbed homes. Performance on these testbed sites provides an indication of how well we can detect activity transitions in general, including the EMA testbed smart homes. Each of the 30 apartments house 1-2 older adult (age 75+) residents who perform daily routines while sensors in the apartment generate and store events. To provide ground truth activity labels, annotators are given the house floor plan, the positions of the sensors, a resident-completed form indicating when and where they typically perform daily activities, and the sequence of sensor events. Multiple annotators provide labels with the beginning and ending of activity occurrences and an inter-annotator agreement of κ=0.80. The activity classes that we use for our analyses are Bathe, Enter Home, Wash Dishes, Personal Hygiene, Relax, Work, Sleep, Leave Home, Cook, Bed Toilet Transition, Eat, and Other Activity. The output of the algorithm is a tag attached to each sensor event indicating the performed activity label.

Finally, we use a semantic reasoner to analyze activity transitions and the activity label tags assigned to sensor events to decide whether time t is an appropriate time to send a prompt. To make the final decision, EMA-SEP verifies that 1) the activity transition tag indicates t is an activity breakpoint, 2) there was no Leave Home activity in the previous 15 minutes (to ensure the participant is home), 3) there was no Sleep activity in the previous 15 minutes (to avoid waking up the participant), and 4) there was no time gap more than an hour in length between two recent sensor events (which may further indicate the participant is out of the home).

B. Experiment Setup

We tested two versions of the SEP-EMA system over two testing iterations in smart home testbeds. In the first iteration, we identified time t as a change point if the corresponding change score is larger than a threshold value. To further reduce the number of false positives, in the second iteration we only considered t as a change point if the corresponding change score not only has a value larger than the threshold but also represents a local maximum. Table II shows a summary of participant demographics and the average number of residents for each iteration. The SEP-EMA algorithm described in Section A is based on the Iteration 2 design. The Iteration 1 design did not include the filtering of double change points.

TABLE II
PARTICIPANTS' DEMOGRAPHICS FOR EACH SEP-EMA ITERATION

	Iteration 1	Iteration 2
Number of participants	N=9	N=10
Mean age in years	65.55	73.70
Gender	9 f, 0 m	8 f, 1 m
Education range (years)	12-20	14-20
Number of residents	2 single, 5 double 2 3+ residents	7 single, 2 double 1 3+ residents

In this study, we installed smart home sensors (including passive infrared motion, and door sensors) in participant homes and collected data while they performed their regular daily activities for three to four months. They received automated information requests via a mobile device for a one-week period on two occasions spaced about one month apart. Each day is segmented into four prompt windows and participants receive one prompt in each window. The window times were typically three-hour periods between 8:00am and 9:30pm with a half hour between each window. One week of prompting was used as a baseline and therefore employed random time prompting. This means a time was randomly selected during each prompt window on each day for EMA questioning. During the other week of prompting, we used SEP-EMA to prompt participants during their activity transitions. In this condition, SEP-EMA selected the first acceptable change point t during each prompt window on each day to ask the EMA questions. To offer a fair comparison between random prompting and transition-aware prompting, we only send random time prompts if the participant is in the home. To control for order effects, the data was

counterbalanced. Half of the participants received random time prompting first and the other half first received prompts based on the SEP-EMA method.

To evaluate the performance of SEP-EMA prompting we analyze participant responses to the prompt timing question. We hypothesize that CPD-based prompting will increase participant response rate over random time prompting and result in higher endorsement of conveniently timed prompts. If the hypothesis is validated, this provides an indication of app usability as well as decreased activity interruption.

III. RESULTS

The results of SEP-EMA prompting depend to a large extent on the ability of the activity recognition algorithm to correctly identify the performed activity and the ability of SEP to correctly identify activity transitions. We therefore evaluate the performance of these two algorithms on data collected from 30 smart homes and analyze the impact of random time and SEP-EMA prompting on response rate and participant satisfaction.

A. Evaluation of Activity Recognition

We evaluate the performance of activity recognition based on 3-fold cross-validation. Cross validation is performed separately for each participant and the results are averaged over the 30 testbed homes. As shown in Figure 3, activity recognition achieves an average of 0.98 accuracy. In addition to accuracy, we also report F1-measure which is calculated as the square root of the product of sensitivity and specificity for a particular class label as shown in (3). Here, TP, FP, TN, and FN refer to true positive, false positive, true negative, and false negative of a specific activity class label. The results are averaged over the classes, providing a macro-average. The F1-measure is particularly valuable when reporting classifier performance on imbalanced datasets. As shown in Figure 3, activity recognition achieved an average F1-measure of 0.98.

$$F1 - Measure = 2\sqrt{\frac{\frac{TP}{TP}\times\frac{TP}{TP+FP}}{\frac{TP}{TP+FN}\times\frac{TP}{TP+FP}}} \quad (3)$$

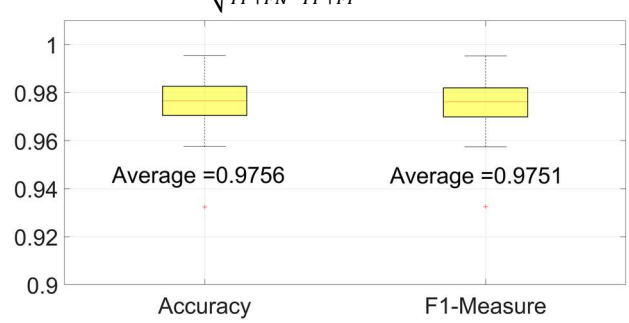

Fig. 3. The accuracy and F1-measure of the AR activity recognition algorithm.

B. Evaluation of Activity Transition Detection

To analyze the performance of our transition detection algorithm, we assume a detected change point is correct if there exists a ground truth change point in the data that occurs soon before or after the detected change point. Here, an "actual" or ground truth change point is based on manually-annotated activity labels. In other words, a detected change point at time t* is correct if an actual change point occurs in the time interval [t* − λ, t* + λ]. In our experiments, we consider λ=5 seconds for evaluation of change point detection with a small time offset. In our experiments, we identified a threshold value of 0.5 that optimizes a tradeoff between the true positive rate (TPR) and false positive rate (FPR) based on a sample of the sensor data.

Figure 4 shows the TPR and FPR for activity transition detection of the 30 testbed homes. These are compared for theSEP, unconstrained least-squares importance fitting (uLSIF) [40], relative unconstrained least-squares importance fitting (RuLSIF) [41], and Bayesian change detection methods based on change points within 10 seconds. Since transition detection is the heart of our prompting management model, its accuracy has a large impact on the response rate. As we can see, SEP outperforms all other methods in detecting activity transitions with an average TPR = 87.93%. The difference is significant at the ($p < .05$) level.

The SEP algorithm exhibits a lower FPR (average = 11.52%) than RuLSIF and uLSIF, but its FPR is higher than the Bayesian algorithm. Recalling that the Bayesian CPD has a very low TPR and thus does not detect changes consistently, we conclude the small FPR value in this method is because of its overall low detection rate. This result indicates that SEP is more robust against noise or outliers than the other compared methods. Because human behavior (and therefore smart home sensor data) is dynamic, SEP performs better on this data and detects fewer false alarms. The one-way ANOVA test indicates the difference between FPR values for SEP and the other algorithms is significant at the ($p < .05$) level. A larger false positive rate has a net effect of splitting the data into more segments. Because some activities such as cooking are complex, splitting the data into more segments likely results in some EMA prompts occurring between phases of these more complex activities. Because activity recognition and change point detection are the foundation of our EMA prompt timing, these results provide support that the change point detection-based prompts will be given close to the actual transition points for individuals in the EMA study.

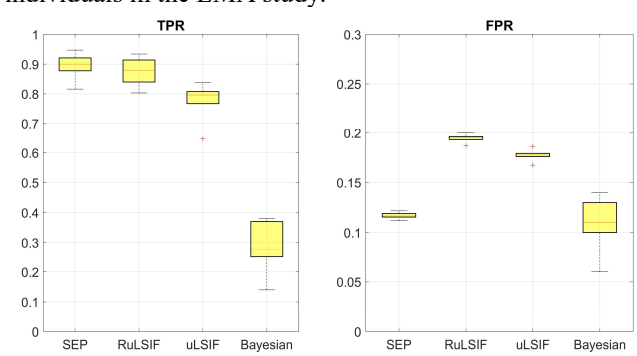

Fig. 4. Activity transition detection TPR and FPR.

C. Evaluation of Prompt Timing

To evaluate SEP-EMA, we calculate user response rate by computing the ratio of the number of answered prompts to the total number of prompts. We begin by comparing the user EMA question response rate for random timing and SEP-EMA notification delivery. A total of 536 prompts were sent across all 19 participants using random timing method and 537 prompts were sent using the SEP-EMA delivery method throughout the study. Figures 5 and 6 show the box plots of the response rates and per home response rate, for both random

timing and SEP-EMA. Using SEP-EMA to deliver notifications to participants increased the average response rate from 78.62±0.34% to 84.26±9.54%. According to a paired t-test, the difference is significant ($p<0.05$). The response rate in ~~16~~ 15 out of 19 homes increased when we used SEP-EMA, but in the remaining ~~3~~ 4 homes the response rate decreased.

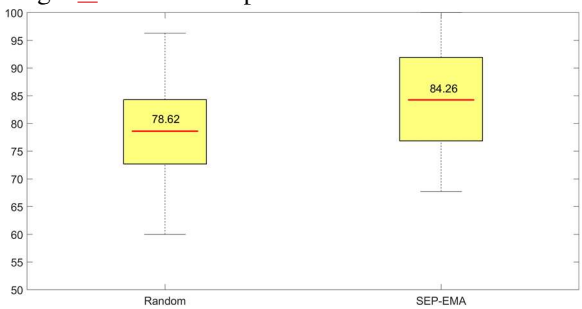

Fig. 5. Response rate of random timing and SEP-EMA for all participants.

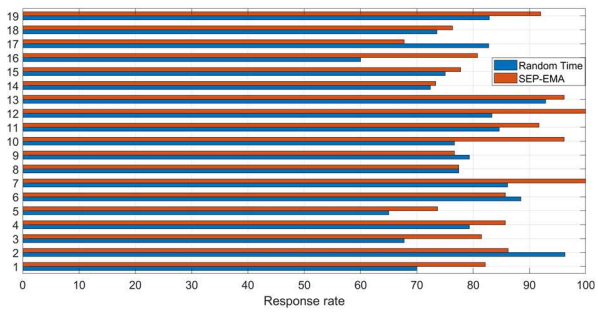

Fig. 6. The response rate of each home summed over all days.

Table III shows the results of 2×2 mixed model ANOVA with delivery method as a within factor, iteration as a between factor, and response rate as a dependent value. The goal of this evaluation is to determine if there is a significant change in response rate due to the changing delivery method and refinement of the algorithm design in two iterations. Based on the first iteration of the EMA app, the average response rate increased from 78.85±10.28% to 83.22±7.70% over random-time prompting. However, based on iteration 2, the average increased from 78.40±8.97% to 85.19±11.29%. As we can see, the effect of notification delivery method is significant ($p<0.05$), but the difference between two iterations of SEP-EMA and the interactions are not significant. In this table, generalized η^2 is a measure of effect size which is the ratio of effect to total variance.

TABLE III
MIXED ANOVA RESULTS TO COMPARE ITERATIONS OF SEP-EMA

	F	P	Generalized η^2
Iteration	0.0011	0.9737	5.22e-05
Delivery method	5.3548	0.0343	7.89e-02
Iteration : Delivery method	0.2086	0.6540	3.33e-03

In addition to evaluating response rates, we investigate the subjects' responses to the question regarding prompt timing. Table IV demonstrates the average percentage of different timing prompt responses for both iterations and in total. The percentage was calculated by dividing the number of answers in each category by the total number of answered prompts. Most participants considered the Neutral and Convenient responses to be similar and used just one of these categories consistently, rather than indicating Neutral on some occasions and Convenient on others. Thus, here we combine Convenient and Neutral answers. Considering all responses, there is no significant difference between random timing and SEP-EMA in terms of percentage of choosing the Convenient/Neutral or Poor answers at $p<0.05$. The most likely explanation is that when prompt timing is not convenient, participants do not answer the prompts at all and we do not include the lack of responses in the totals. Thus, one alternative way to calculating response rate would be considering the Poor timing answer as lack of answer. In this case, similar to previous result, the average of subjective responses increased from 73.61±13.69% to 79.99±12.39% when we use SEP-EMA instead of random timing.

TABLE IV
THE AVERAGE PERCENTAGE OF DIFFERENT TIMING PROMPT RESPONSE

	Convenient/Neutral		Poor	
	Random	SEP-EMA	Random	SEP-EMA
Iteration 1	96.79	95.06	3.21	4.94
Iteration 2	90.39	94.51	9.61	5.49
Total	93.43	94.77	6.57	5.23

Next, we perform an analysis to understand the role that a person's current activity plays in responding to an EMA prompt. Specifically, we calculate the number of total prompts and number of answered prompts for each automatically-detected activity. In the case of random-timing prompts, the detected activity is the one that is occurring at the time of the prompt. For SEP-based prompt timing, the prompt occurs between activities. As a result, we look two time periods: 1) five minutes before the prompt and 2) five minutes after the prompt response. The most commonly-occurring activity within this combined time period is used as the detected activity.

Aligning detected activities with EMA responses is more challenging in multi-resident homes because the activity recognition algorithm does not identify the specific resident that is performing the activity. To eliminate the impact of multi-resident homes, we only consider the nine single-resident homes in this study. Figure 7 shows the overall per-activity response rate. As we can see there is a statistically-significant relationship between current activity and response rate ($p<0.01$). Activities Cook and Relax yield the highest response rates. The Cook activity is typically complex with multiple subtasks, so one explanation for the increased response rate here is that it is easier for residents to answer the notifications during the transitions between phases of the activity. In the case of Relax, the activity is not critical, so interruptions are likely not disruptive. On the other hand, no participant responded to any prompt during Enter Home or Wash dishes which is logical due to the nature of these activities.

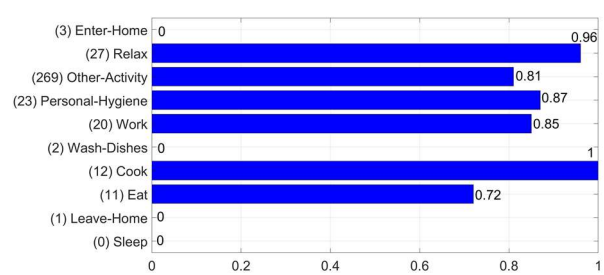

Fig. 7. Overall per-activity response rate for single-resident homes. Numbers before activity labels indicate the number of times individuals were prompted when these activities occurred.

IV. DISCUSSION

Advances in smart environments have enhanced the development of EMA by enabling them to collect more accurate data at desired times. Although app-based EMA is becoming ubiquitous, current approaches employ rule-based methods to time the sending of prompts. The present study investigated the direct relationship between EMA response rates and prompt delivery methods. We proposed a new EMA method that utilizes activity recognition and change point detection to provide prompt timings. We hypothesized that such context-aware EMA delivery would increase the response rate of EMA questions in comparison with a random-time method.

Consistent with our hypothesis, transition-based context-aware delivery of EMA notifications did offer benefits for increasing response rate compared to the random time-based delivery - participants were more likely to respond to prompts sent based on SEP-EMA timing than based on random timing. As mentioned earlier, in existing EMA systems, researchers often use additional incentives to encourage the participant to answer questions to reach the minimum 75% or 80% response rate. For example, Fritz et al. [42] assigned an assistant to each participant who was responsible for monitoring the real-time data collection, contacting the participant if they did not respond, and conducting additional troubleshooting and training as needed. On the other hand, Maher et al. [43] paid $80 to participants to answer at least 80% of the EMA prompts. In contrast, SEP-EMA alone, without additional assistance, increases response rates for EMA data collection by almost 7%. Only 68% of participants demonstrated a response rate greater than the threshold based on randomly-timed prompts (75%), but after using SEP-EMA, 100% of the participants demonstrated response rates over 75%.

While earlier work revealed the connection between a person's current activity and their interruptibility [28][29], in this study, we analyze this relationship in terms of automatically-detected activities and actual EMA responses. The current study shows participants are more willing to answer prompts when they are performing complex multistep activities like Cooking because there may be natural breaking points within the activity itself. We also observe that residents are more willing to respond to prompts when they are in the middle of activities that appear to be non-critical like Relax. However, further investigation is required to understand the relationship between response rates and Work, Personal Hygiene, and Eat.

Although our experimental results show that a transition-based context-aware prompting system can be useful in many ways, there are some limitations in the study that need to be addressed. Data collection in the current study was limited to only 19 participants and these individuals are midlife or older adults, therefore we do not know how well transition-based prompting works for younger individuals. Since the younger generation are regular mobile phone users and spend more time near their smart phones, prompting them during transitions between activities may not provoke as large a difference as for older adults. The future work could replicate this study with a larger study population for a longer study duration.

Although the current experiment was performed in partnership with a smart home platform and activity transitions and labels were identified using smart home sensor data, the proposed method can be implemented using virtually any type of activity monitoring sensor platform, such as smartphone or smartwatch. Delivering SEP-EMA to an individual's mobile device allows the EMA platform to assess phenomenon of interest at any location, both inside and outside the home. It would also avoid the difficulties that accompany activity monitoring in smart homes with multiple residents.

Finally, effective EMA delivery needs to consider be effective for multiple types of sensor data. Because smart home systems can cost hundreds or thousands of dollars, the usability may be limited to those who can afford the technology and the increase in response rate may not consistently justify the cost of installing smart home sensors. We are currently enhancing SEP-EMA to process multiple types of sensor data, which will generalize its benefit to mobile platforms as well as smart home settings.

V. CONCLUSIONS

In this paper, we introduce a context-aware algorithm, SEP-EMA, to improve the performance of EMA in managing prompts and notifications. SEP-EMA combines traditional EMA technology with knowledge about a person's current activity, activity transitions, and other contextual information to find the best time to ask EMA questions. Experimental results, obtained from a set of participants answering EMA queries while in their smart homes, indicate that the proposed algorithm outperforms traditional random time-based prompting. We therefore conclude that using SEP-EMA increases response rates and decreases activity interruptions. The resulting approach is useful for presenting information, interacting with individuals without creating interruption overload, and boosting EMA data collection process.

REFERENCES

[1] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological Momentary Assessment," *Annu. Rev. Clin. Psychol.*, vol. 4, no. 1, pp. 1–32, Apr. 2008.

[2] J. D. Runyan and E. G. Steinke, "Virtues, ecological momentary assessment/intervention and smartphone technology," *Front. Psychol.*, vol. 6, p. 481, May 2015.

[3] D. J. Beal and H. M. Weiss, "Methods of Ecological Momentary Assessment in Organizational Research," *Organ. Res. Methods*, vol. 6, no. 4, pp. 440–464, Oct. 2003.

[4] M. S. Mundi, P. A. Lorentz, K. Grothe, T. A. Kellogg, and M. L. Collazo-Clavell, "Feasibility of Smartphone-Based Education Modules and Ecological Momentary Assessment/Intervention in Pre-bariatric Surgery Patients," *Obes. Surg.*, vol. 25, no. 10, pp. 1875–1881, Oct. 2015.

[5] L. L. Willett *et al.*, "Use of Ecological Momentary Assessment to Guide Curricular Change in Graduate Medical Education," *J. Grad. Med. Educ.*, vol. 3, no. 2, pp. 162–167, Jun. 2011.

[6] T. M. Piasecki *et al.*, "The subjective effects of alcohol-tobacco co-use: an ecological momentary assessment investigation.," *J. Abnorm. Psychol.*, vol. 120, no. 3, pp. 557–71, Aug. 2011.

[7] E. M. Kleiman, B. J. Turner, S. Fedor, E. E. Beale, J. C. Huffman, and M. K. Nock, "Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies.," *J. Abnorm. Psychol.*, vol. 126, no. 6, pp. 726–738, Aug. 2017.

[8] J. P. Maher, A. L. Rebar, and G. F. Dunton, "Ecological Momentary Assessment Is a Feasible and Valid Methodological Tool to Measure Older Adults' Physical Activity and Sedentary Behavior.," *Front. Psychol.*, vol. 9, p. 1485, 2018.

[9] T. Okoshi, J. Ramos, H. Nozaki, J. Nakazawa, A. K. Dey, and H. Tokuda, "Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones," in *2015 IEEE International*

Conference on Pervasive Computing and Communications (PerCom), 2015, pp. 96–104.

[10] V. Pejovic and M. Musolesi, "InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*, 2014, pp. 897–908.

[11] L. G. Morrison *et al.*, "The Effect of Timing and Frequency of Push Notifications on Usage of a Smartphone-Based Stress Management Intervention: An Exploratory Trial," *PLoS One*, vol. 12, no. 1, p. e0169162, Jan. 2017.

[12] M. H. Vastenburg, D. V. Keyson, and H. de Ridder, "Considerate home notification systems: a field study of acceptability of notifications in the home," *Pers. Ubiquitous Comput.*, vol. 12, no. 8, pp. 555–566, Nov. 2008.

[13] D. S. Moskowitz and S. N. Young, "Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology.," *J. Psychiatry Neurosci.*, vol. 31, no. 1, pp. 13–20, Jan. 2006.

[14] J. D. Runyan, T. A. Steenbergh, C. Bainbridge, D. A. Daugherty, L. Oke, and B. N. Fry, "A Smartphone Ecological Momentary Assessment/Intervention 'App' for Collecting Real-Time Data and Promoting Self-Awareness," *PLoS One*, vol. 8, no. 8, Aug. 2013.

[15] P. van de Ven *et al.*, "ULTEMAT: A mobile framework for smart ecological momentary assessments and interventions," *Internet Interv.*, vol. 9, pp. 74–81, Sep. 2017.

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

[17] K. M. Miller, C. C. Price, M. S. Okun, H. Montijo, and D. Bowers, "Is the n-back task a valid neuropsychological measure for assessing working memory?," *Arch. Clin. Neuropsychol.*, vol. 24, no. 7, pp. 711–7, 2009.

[18] W. van Ballegooijen, J. Ruwaard, E. Karyotaki, D. D. Ebert, J. H. Smit, and H. Riper, "Reactivity to smartphone-based ecological momentary assessment of depressive symptoms (MoodMonitor): protocol of a randomised controlled trial.," *BMC Psychiatry*, vol. 16, no. 1, p. 359, 2016.

[19] M. F. Armey, H. T. Schatten, N. Haradhvala, and I. W. Miller, "Ecological momentary assessment (EMA) of depression-related phenomena," *Curr. Opin. Psychol.*, vol. 4, pp. 21–25, Aug. 2015.

[20] V. M. Gonzalez and P. L. Dulin, "Comparison of a smartphone app for alcohol use disorders with an Internet-based intervention plus bibliotherapy: A pilot study.," *J. Consult. Clin. Psychol.*, vol. 83, no. 2, pp. 335–45, Apr. 2015.

[21] J. Morgenstern, A. Kuerbis, and F. Muench, "Ecological Momentary Assessment and Alcohol Use Disorder Treatment.," *Alcohol Res.*, vol. 36, no. 1, pp. 101–9, 2014.

[22] G. Knell, K. P. Gabriel, M. S. Businelle, K. Shuval, D. W. Wetter, and D. E. Kendzor, "Ecological Momentary Assessment of Physical Activity: Validation Study.," *J. Med. Internet Res.*, vol. 19, no. 7, p. e253, Jul. 2017.

[23] G. F. Dunton, "Ecological Momentary Assessment in Physical Activity Research.," *Exerc. Sport Sci. Rev.*, vol. 45, no. 1, pp. 48–54, 2017.

[24] K. L. Watkins *et al.*, "Advancing Cessation Research by Integrating EMA and Geospatial Methodologies: Associations Between Tobacco Retail Outlets and Real-time Smoking Urges During a Quit Attempt," *Nicotine Tob. Res.*, vol. 16, no. Suppl 2, pp. S93–S101, May 2014.

[25] L. K. Dhingra *et al.*, "Ecological Momentary Assessment of Smoking Behavior in Persistent Pain Patients," *Clin. J. Pain*, vol. 30, no. 3, pp. 205–213, Mar. 2014.

[26] B. P. Bailey and J. A. Konstan, "On the need for attention award systems: Measuring effects of interruption on task performance, error rate, and affective state," *J. Comput. Hum. Behav.,* vol. 22, no. 4, pp. 709–732, 2006.

[27] A. Mehrotra and M. Musolesi, "Intelligent Notification Systems: A Survey of the State of the Art and Research Challenges," Nov. 2017.

[28] T. Gillie and D. Boradbent, "What makes interruptions disruptive? A study of length, similarity, and complexity," *Psychol. Res.*, vol. 50, no. 4, pp. 243–250, 1989.

[29] Y. Miyata and D. A. Norman, "Psychological issues in support of multiple activities," in *User Centered System Design: New Perspectives on Human-Computer Interaction*, D. A. Norman and S. W. Draper, Eds. Boca Raton, FL: CRC Press, 1986, pp. 265–284.

[30] S. Rosenthal, A. K. Dey, and M. Veloso, "Using Decision-Theoretic Experience Sampling to Build Personalized Mobile Phone Interruption Models," in *International Conference on Pervasive Computing*, 2011, pp. 170–187.

[31] S. Azizi *et al.*, "Deep Recurrent Neural Networks for Prostate Cancer Detection: Analysis of Temporal Enhanced Ultrasound," *IEEE Trans. Med. Imaging*, vol. 37, no. 12, pp. 2695–2703, Dec. 2018.

[32] A. L. Beam and I. S. Kohane, "Big Data and Machine Learning in Health Care," *JAMA*, vol. 319, no. 13, p. 1317, Apr. 2018.

[33] S. Aminikhanghahi, R. Fallahzadeh, M. Sawyer, D. J. Cook, and L. B. Holder, "Thyme: Improving smartphone prompt timing through activity awareness," in *Proceedings - 16th IEEE International Conference on Machine Learning and Applications*, 2017.

[34] A. Exler, M. Braith, A. Schankin, and M. Beigl, "Preliminary investigations about interruptibility of smartphone users at specific place types," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 1590–1595.

[35] J. E. Fischer, C. Greenhalgh, and S. Benford, "Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications," in *International Conference on Human Computer Interaction with Mobile Devices and Services*, 2011, pp. 189–190.

[36] T. Okoshi, H. Nozaki, J. Nakazawa, H. Tokuda, J. Ramos, and A. K. Dey, "Towards attention-aware adaptive notification on smart phones," *Pervasive Mob. Comput.*, vol. 26, pp. 17–34, Feb. 2016.

[37] C. Park, J. Lim, J. Kim, S.-J. Lee, and D. Lee, "Don't Bother Me. I'm Socializing!," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2017, pp. 541–554.

[38] S. Aminikhanghahi, T. Wang, D. J. Cook, and I. Fellow, "Real-Time Change Point Detection with application to Smart Home Time Series Data," *IEEE Trans. Knowl. Data Eng.*, 2018.

[39] D. J. Cook, N. C. Krishnan, and P. Rashidi, "Activity Discovery and Activity Recognition: A New Partnership," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 820–828, Jun. 2013.

[40] T. Kanamori, S. Hido, and M. Sugiyama, "A Least-squares Approach to Direct Importance Estimation," *J. Mach. Learn. Res.*, vol. 10, pp. 1391–1445, 2009.

[41] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-point detection in time-series data by relative density-ratio estimation.," *Neural Netw.*, vol. 43, pp. 72–83, Jul. 2013.

[42] H. Fritz, W. Tarraf, D. J. Saleh, and M. P. Cutchin, "Using a Smartphone-Based Ecological Momentary Assessment Protocol With Community Dwelling Older African Americans," *Journals Gerontol. Ser. B Psychol. Sci. Soc. Sci.*, vol. 72, no. 5, p. gbw166, Jan. 2017.

[43] J. P. Maher, A. L. Rebar, and G. F. Dunton, "Ecological Momentary Assessment Is a Feasible and Valid Methodological Tool to Measure Older Adults' Physical Activity and Sedentary Behavior," *Front. Psychol.*, vol. 9, p. 1485, Aug. 2018.