Combining Participatory and ESM: A Hybrid Approach to Collecting Annotated Mobility Data

Hsiu-Chi Chang
National Chiao Tung University
Hsinchu, 30010, TW
wbest601.cs05g@g2.nctu.edu.tw

Yung-Ju Chang
National Chiao Tung University
Hsinchu, 30010, TW
armuro@cs.nctu.edu.tw

Mark W. Newman
University of Michigan
Ann Arbor, Michigan, USA
mwnewman@umich.edu

Chih-Hsin Lin
National Chiao Tung University
Hsinchu, 30010, TW
neu.cs07g@nctu.edu.tw

CHI'20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA

ACM ISBN 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.3383066

Abstract
Collecting continual labeled activity data entails considerable effort from users to label a series of activity data. We propose Checkpoint-and-Remind (CAR), a hybrid approach that combines participatory sensing (PART) and context-trigger ESM labeling (ESM). Checkpoint-and-Remind has the advantage of user control and reduces users' burden in recording activities. Meanwhile, it features a context-trigger mechanism as a backup to remind users of labeling. Our preliminary evaluation of CAR with nine participants, who collected and labeled their mobility activity data for 15 weekdays, showed that compared with PART and ESM, participants collected a larger amount of annotated mobility data using CAR. In addition, participants had a higher annotation rate when using CAR than when using ESM. Our results showed that this hybrid approach that combines manual and automated recording is promising. Our future work is validating these results and measure more metrics related to compliance with more participants.

Author Keywords
Annotation; label; ground truth; activity collection; transportation; field experiment; wearable camera

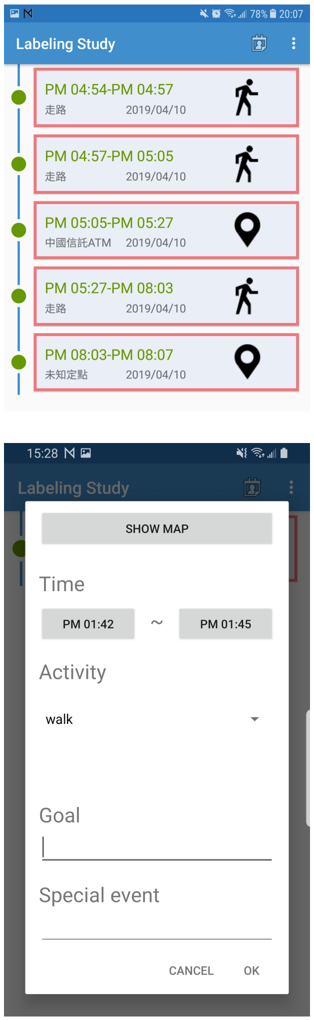

Figure 1: (a)Timeline page, (b) the dialog prompted from clicking each record to fill in the annotations.

CSS Concepts

- **Human-centered computing~Human computer interaction (HCI)**

Introduction

Facilitating the collection of labeled activity data via crowdsourcing has been one of the important topics in mobile crowdsourcing. It is because it is vital to obtain input from users to improve the performance of context recognition. Yet, getting inputs from users pose the burden of data collection and labeling on them. Many researchers have started looking at various approaches to collecting users' inputs. For example, [4] motivated users to provide their data by involving gamification and social incentives. [6] used a powerful label-reporting interface to engage users with different behavior styles and phone-interaction preferences and to acquire detailed labels. [7] investigated the use of situated devices to collect labels in the home environment. [5] developed a video-based annotation tool for parents to capture and annotate in-home problem behaviors of children. [3] used a speech-based approach for labeling sensor data.

Other researchers have also attempted to compare different approaches to collecting labeled activity data. For example, using smartphone as the major tool for labeling travel activity data, [1] showed that using a participatory approach (PART), i.e. users manually start and stop a recording, led to more accurate recording of activity data compared to a context-triggered experience sampling method (ESM) that prompted users to label the activity at the moment detected by the system [1,2]. However, it was also shown that the participatory approach made participants feel more burden and led to fewest recordings. Considering both user control and burden, beyond the methods compared in [1], we propose a hybrid method that combines a participatory and context-triggered experience sampling method called *Checkpoint-and-Remind* (CAR) to collect mobility activity data. Specifically, using CAR, the user checkpoints at transitions between two activities to indicate activity switching. Meanwhile, the system uses context-triggered ESM as a back-up mechanism to record the activity if the user does not checkpoint and to remind the user of labeling the activity.

To evaluate the effectiveness of CAR, we conducted a preliminary field evaluation with nine participants, who recorded and collected their mobility activity data using all the PART, ESM, and CAR methods respectively in 15 weekdays, each method for five days. We found that, overall, compared with PART and ESM, participants collected a larger amount of annotated mobility data using CAR. Participants had a higher annotation rate when using CAR than when using ESM. These show that the hybrid approach that combines manual and automated recording is promising.

The Research App for Data Collection

We developed a research android app that users can install on their phones to record and annotate their mobility activity data. There are two main parts of the app: annotation and recording. For the former, as shown in Figure 1a, the app shows the timeline of the user's mobility activity, including *on foot*, *on the bicycle*, *in the vehicle* and *static*, which are recorded by the app. The former three activities are travel activities, i.e. users traveling from one place to another in a particular transportation mode, whereas the *static* activity means users dwelling at a particular place. To annotate an

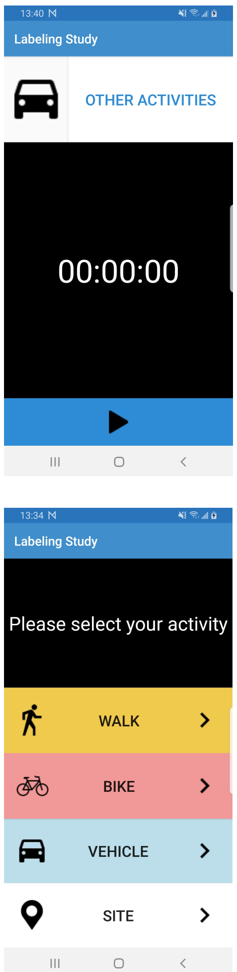

Figure 2: (a) participatory (PART), activity selecting page, (b) participatory (PART), recording page,

activity recording, users choose an activity to label and then pick an activity label from a dropdown menu (Figure 1b). Users can also enter the goal and note special events during the activity. Particularly for labeling static activity, users need to choose a place on a map. If the place has not yet existed on the map, they can create the place. Activity recordings are created when users use either of the three methods to record their mobility activity.

Collecting mobility activity data by PART, ESM, CAR
Similar to the mechanism in [1], when using PART, users start and stop the recording of their activities by themselves (Figure 2a). Before recording, users first choose the type of activity (i.e. labeling them), as shown in Figure 2b.

When using ESM, the app detects the transportation activity of the users. The app builds its transportation detection on top of the results from the Google Activity Recognition [8] result. Inspired by [1], it uses a finite-state machine to determine a transportation mode label. When detecting a new transportation mode, the app records the activity. When it detects that the current activity has started, it notifies the users to that they have a new recording to label (Figure 3a).

When using CAR, users choose a time point when they think a transition between two activities happens, and then "checkpoint" at that time point by pressing the checkpoint button (as shown in Figure 3b). This action makes the app switch the activity (i.e. stop the recording of the current activity and start the recording of the next one). It also informs the app that it should start to detect a new different activity because the current one has ended. If the app finds that the users have switched their activity before they checkpoint, it first waits for the users to checkpoint for one minute. If the users do not checkpoint in a minute, the app itself automatically records the current activity and then, using the ESM approach, reminds the users of labeling it. Below we describe our preliminary field evaluation.

The Field Evaluation

Study Design and Participants
We adopted a within-subject design. Each participant was assigned to a random and counterbalanced order of the three conditions PART, ESM, and CAR to use these methods to record and label their mobility data respectively for five weekdays, thus in total 15 weekdays. We recruited participants from two major social media platforms in Taiwan: Facebook and PTT (a Bulletin Board System in Taiwan). Our selection criterion was that the participants had to regularly commute so that they could collect their data routinely. All the study participants regularly commuted by with at least kinds of transportation modes on weekdays.

Study Procedure

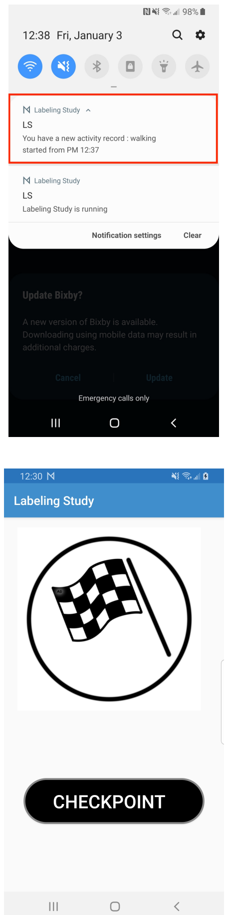

Figure 3: (a) context-triggered experience sampling method (ESM), notification reminded users to label and annotate, (b) Checkpoint-and-Remind (CAR), checkpoint button. Press to tell the instrument user's activity just being switched.

We asked participants to collect data on weekdays because we assumed that mobility on weekends might be more unexpected. We asked them to try their best to record and label their mobility activities for each entire day during the study. We asked them to choose a label "static" whenever they thought they were not moving but staying at a specific place. In addition to labeling, we also asked them to annotate each travel activity with the goal of that trip. A "special event" field was provided but optional. For this, we encouraged them to note anything they thought that made that activity different from their regular routine activity. For example, if they encounter a car accident during the trip, then they can fill in "encounter a car accident". Each recording was shown on the timeline page of the app. At midnight, the page was refreshed in order to display the daily mobility timeline in the next day. In addition to using the app to collect mobility data, we also asked participants to wear a wearable camera provided by us when they were traveling outdoors. The camera took a photo every 10 seconds, and we used these photos to help us reconstruct the ground truth of the participants' mobility history during the study period. With these photos, we could compare which method allowed participants to record and label more mobility data. This approach was also adopted by [1].

During the study period, participants were asked to send their activity data and the photos from the wearable camera every day. They could review and remove any photos they were reluctant to share with the research team. After completing the use of each data collection approach, participants filled out a questionnaire to measure their perceived effort of using the method. Upon the completion of the 15-weekday data collection, participants were provided $1200 NTD (roughly 40 USD) and invited to a post-study interview. If they attended the interview, they obtained an additional $300 NTD.

Data Processing and Analysis

Ground Truth Trips Generating

To compare the mobility data collected using the three methods, we established the ground truth of their mobility data from the photos of the wearable camera and the GPS logs collected from the phone. Two authors of the paper coded the start and the end times of each mobility activity based on the photos and the GPS races replayed on Google Earth for Desktop [9]. They first coded a subset (5.7%) of the dataset and iteratively discuss the recognition of the start and the end of activity till they reached an intra-class coefficient (ICC) score of 0.855, indicating high reliability between the two coders. After the test, they coded the rest of the photos. We generated 1,469 ground truth mobility activity instances and paired each of them with the activity recording collected by the participants.

We ran the mixed-effect regression model for examining the effect of collection approaches on the amount of activity data collected. We also analyzed the labeling and annotation of coverage using each collection approach. We define that an activity recording was labeled only if the user had chosen an activity label and annotated only if the user had noted the goal and chosen an activity label.

Results

We collected 2,735 recordings and reconstructed 1,469 ground truth activities from the nine participants. After combining and cleaning erroneous recordings, we examined 1,455 recordings with the ground truth trips.

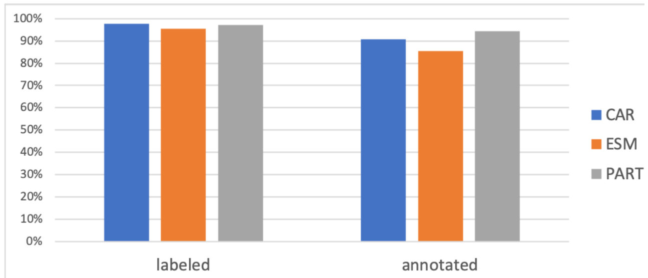

Figure 4: the labeled and annotated rate.

We ran a mixed-effect logistic regression to examine whether a trip was recorded and labeled/annotated, respectively. We included condition, the order of the assigned conditions, and the ground truth activity (static, walk, bike, driver, passenger) in the model to examine their main effect on the outcome variables, which included two binary variables of labeling and annotating (either true or false), and the amount of labeled and annotated data daily, respectively. We analyzed the transcripts obtained from the post-study interview using affinity diagrams.

Label & Annotations Compliance

We first examine participants' compliance with labeling and annotation at the individual-recording level. That is, for each mobility recording, how often did participants label and annotate the recording, respectively? Figure 4 (Left) shows that the labeling rates in the three conditions were all very high: 97.17% (PART), 95.59% (ESM), and is 97.58% (CAR). However, among the three conditions, participants were the least likely to label a recording in the ESM condition, and we did not observe a significant difference between PART and CAR (PART vs. ESM: $Z = -1.735$, $p = 0.083$; ESM vs. CAR: $Z = 2.260$, $p = 0.024$; PART vs. CAR: $Z = 0.505$, $p = 0.613$).

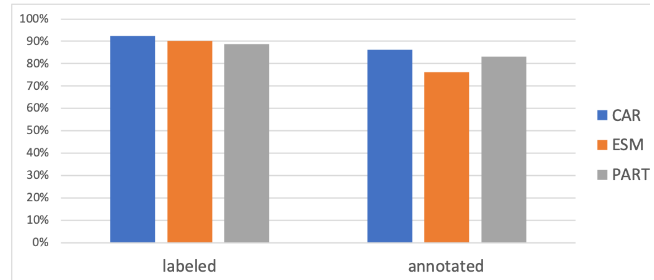

Figure 5: average coverage per day by percentage.

On the other hand, participants were more likely to annotate (Figure 4 Right), i.e. providing at least the goals for the mobility recordings, in the PART condition (94.35%) than in the other two conditions. This was likely that in the PART condition, the research apps generated fewer recordings for participants to annotate (n=460), as indicated in [1]. Interestingly, with a similar number of mobility recordings between the ESM and CAR conditions, participants were also more likely to annotate in the CAR condition than in the ESM condition (CAR: 90.73%, n=496; ESM: 85.37%; n=499). All of the differences were statistically significant (PART vs. ESM: $Z = -3.998$, $p < 0.001$; ESM vs. CAR: $Z = 2.186$, $p = 0.029$; PART vs. CAR: $Z = -2.020$, $p = 0.043$). These results suggest that with more user involvement in determining the timing of mobility recording (PART, CAR), participants had higher compliance for providing additional annotations than when there was no user involvement (ESM). Our qualitative findings also indicate that participants preferred more user involvement than not any.

Amount of Labeled Data

Next, we examined the amount of labeled mobility data at a day-level. We measured the amount of data using a notion of *coverage*, which was measured in terms of time (in seconds), adopted from [1]. The data coverage was compared with the length of ground truth trips we reconstructed earlier. That is, for each day, we asked: how long were entire-day mobility data were added a label and annotation respectively by the participants.

We found that participants' labels achieved similar data coverage across the three conditions. All of them were roughly 90% (PART: 88.7%, ESM: 90%; CAR: 92.3%, Figure 5 Left), suggesting that among the mobility data captured by the wearable camera, most were added a label. However, the coverage of their annotations in the ESM condition was noticeably lower than the coverage of their annotations in the other two conditions (PART: 83.1%, ESM: 76.1%; CAR: 86.2% Figure 5, Right). Although we did not observe a statistically significant difference (PART vs. ESM: $t = -1.42$, $p = 0.158$; ESM vs. CAR: $t = 0.565$, $p = 0.573$; PART vs. CAR: $t = -0.861$, $p = 0.391$), we think it was because the sample size was small. However, this result is promising, because it suggests that participants contributed the most amount of annotated mobility recordings using CAR, in terms of time, throughout a day among the three approaches.

Finally, our qualitative findings suggest that participants felt that PART produced the most accurate data, which was consistent with [1]. But they also complained that when they were in a hurry, it was hard to start or end a trip. In contrast, the fact that CAR only required them to checkpoint at the moment of activity-switching dramatically reduced their burden. Furthermore, they felt less pressure when using CAR than when using PART because they were not worried about not recording their mobility since CAR would record for them when they forgot or when they were too busy to record. The majority of the participants (4/8) preferred CAR due to the moderate controllability. The rest preferred either of the other two for other reasons. In terms of perceived effort, participants still thought the ESM introduced the least effort among the three approaches ($M=3.5$), and felt that CAR($M=4.63$) took less effort than PART did (5.38).

Conclusion:

Our results show that CAR led to larger coverage of annotated data than ESM. However, ESM did not lead to larger coverage of labeled data than CAR did. This shows that the hybrid method seemed to outperform both the participatory approach and the full automation approach. Although participants had the highest annotation rate when using PART, it did not result in the largest coverage of mobility data to which participants added annotations. In addition, the participants also generally agreed that PART took the most effort than the other two conditions. These preliminary results show that CAR is a promising approach. Although it takes less effort, it did not sacrifice the quantity of labeled and annotated data. Our future work is to recruit more participants to validate this preliminary conclusion, as well as compare more compliance metrics among the three methods, such as how many labels were labeled manually versus automatically by the app, as well as participants' perceived workload.

Reference:

[1] Yung-Ju Chang, Gaurav Paruthi, Hsin-Ying Wu, Hsin-Yu Lin, and Mark W. Newman. 2017. An

investigation of using mobile and situated crowdsourcing to collect annotated travel activity data in real-word settings. International Journal of Human-Computer Studies 102: 81–102.

[2] Ian Cleland, Manhyung Han, Chris Nugent, et al. 2014. Evaluation of Prompted Annotation of Activity Data Recorded from a Smart Phone. Sensors 14, 9: 15861–15879.

[3] Susumu Harada, Jonathan Lester, Kayur Patel, et al. 2008. VoiceLabel: using speech to label mobile sensor data. Proceedings of the 10th international conference on Multimodal interfaces, ACM, 69–76.

[4] Theus Hossmann, Christos Efstratiou, and Cecilia Mascolo. 2012. Collecting Big Datasets of Human Activity One Checkin at a Time. Proceedings of the 4th ACM International Workshop on Hot Topics in Planet-scale Measurement, ACM, 15–20.

[5] N. Nazneen, Agata Rozga, Mario Romero, et al. 2012. Supporting parents for in-home capture of problem behaviors of children with developmental disabilities. Personal and Ubiquitous Computing 16, 2: 193–207.

[6] Yonatan Vaizman, Katherine Ellis, Gert Lanckriet, and Nadir Weibel. 2018. ExtraSensory App: Data Collection In-the-Wild with Rich User Interface to Self-Report Behavior. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, ACM, 554:1–554:12.

[7] Heed: Exploring the Design of Situated Self-Reporting Devices: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies: Vol 2, No 3. Retrieved January 6, 2020 from https://dl.acm.org/doi/10.1145/3264942.

[8] ActivityRecognitionResult | Google APIs for Android. Retrieved January 5, 2020 from https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionResult.

[9] Google Earth. Retrieved January 1, 2020 from https://earth.google.com/web/.