The Application of Smartphones in Social Sciences Studies

R.J. Klaasse Bos

December 17, 2017

Executive Summary

Prior research has demonstrated that self-report data from respondents in experience sampling research studies is far from accurate. A new approach, known as reality mining, using smartphones to track participants' location and usage statistics enables accurate, cost-efficient, large-scale and unobtrusive data-collection for research purposes. Thus far, scientists have developed in-house custom mobile applications as there is no universal reality mining tool on the market yet. This business case proposes a feature set for such a mobile application based on suggestions in scientific literature and discusses the project's scope, competitive advantage and financial feasibility.

Contents

1 Opportunity

1.1 Problem

In 40% of mobile phone use studies published in communication journals the empirical analysis is based on self-report data from respondents (Boase & Ling, 2013) [2]. The same study shows that self-report data correlates only moderately with server log data. The authors mention underlying causes that can account for this difference. On the one hand, respondents may find it difficult to estimate the frequency and duration of their own usage behaviour, especially on a continuous scale. On the other hand, respondents tend to under- or overestimate their usage because of social desirability.

Given these findings, the authors finally conclude that there is good reason to be suspicious towards psychological research that claims to have found significant correlations between self-report data and other variables. This conclusion is shared by a more recent study by Andrews et al. (2015) [1] in which they compared 23 participants actual smartphone use with self-reported estimates and found that the actual number of daily pickups was twice as high as what respondents reported. For that reason Boase and Ling recommend to use a more accurate source of data, for example a mobile application which directly sends log data from smartphones to researchers.

1.2 Solution

As follows from the problem statement researchers have suggested smartphones as a more robust alternative for data collection. This is also known as reality mining: automatic data collection that does not require any questionnaire to be filled out or any deliberate user interaction.

In section 1.2.1 the arguments for choosing a smartphone and the current challenges thereof are discussed based on the existing body of literature. After that, in section 1.2.2 a concrete reality mining application is proposed in accordance with scholars' suggestions and recommendations. Thereafter, the product scope and roadmap are defined (1.2.3) and compared to alternative solutions currently available on the market (1.2.4) and finally the customer segments are specified (1.2.5).

1.2.1 Suggestions in Literature

A study back in 2009 by Raento et al. [9] already argued that a smartphone is a useful tool for social sciences research because it allows for cost-effective access to previously inaccessible sources of data. The main arguments for using a smartphone in research are:

- *Adoption rate:* In developed countries a large fraction of people own a phone which already forms an integral part of their daily life. That is to say, respondents do not have to charge and bring along another device throughout the day. In other words, the willingness of people to carry phones is a key argument.
- *Willingness to participate:* Based on their own experience the researchers found that it was not difficult to persuade someone to switch their smartphone in research setting as long as the data was being tracked for research purposes. A study by the The Netherlands Institute for Social Research (2013) [10] even found that respondents would not be concerned if their GPS location was being tracked without knowing it, given that it would be used for a research study.
- *Cost-efficiency:* Using a smartphone as a method of data collection enables medium-longterm research on a large scale thanks to its low fixed and variable costs.
- *Unobtrusive data collection:* Contrary to an observational study the researcher does not directly influence the respondent since the mobile application is running in the background.
- *Correctness data:* Alternatives used in experience sampling - diaries, interviews and questionnaires - have typical threats of validity while the reliability of smartphones and cellular connections have become more mature over time.

Although the automation of observation greatly reduces the amount of work for researchers, Raento et al. (2009) [9] and Sonck and Fernee (2013) [10] acknowledge that there are still some challenges to overcome:

- *Technical:* Programming additional features can be a daunting task, tools to support the analysis of data gathered with smartphones are not widely available and representation and visualisation of high volumes of multidimensional timeline data remains a difficult task.
- *Missing data:* Participants may not always take their phone with them (e.g. they leave their phone at the office during the break when they go out for lunch).
- *Representativeness*: Not everybody can afford or wants to use a mobile phone which can lead to selection and coverage biases if the outcomes of the study are intended to be representative for the general population.
- *Ethical*: Even though researchers explicitly state what data will be collected it is often unsure for participants what insights and patterns can be gained with it.

Despite these drawbacks the authors of both studies believe that smartphones can be used as a main or supplementary tool for many experience sampling studies. Even more, Hofmann and Patel (2015) [7] expect that the experience sampling approach becomes an increasingly popular tool in the years to come thanks to the increasing feasibility and ease of use.

1.2.2 Value Proposition

The main advantage of reality mining is that it can replace survey questions and that the data obtained is generally more accurate compared to self-reported estimates of behaviour. Yet, there is no off-the-shelf universal tool available that investigators can employ to track and gather mobile usage and location data among participants for research purposes. For that reason researchers have developed custom software specific to their needs thus far (e.g. the studies by Sonck and Fernee [10]; Raento et al. [9]; Doherty et al. [4]). These practices are not only costly and time-consuming, they also do not benefit the longevity of the custom software due to the rapid developments in the mobile phone market which demands an ongoing and recurring development effort.

Hence, *an universal reality mining app is developed that tracks participant's everyday mobility patterns, mobile activity throughout the day and device status and automatically transfers the collected data to researchers in a clean format*. Further, it supports the invitation, registration and screening of research participants (Appendix B).

1.2.3 Scope and Boundaries

To avoid sample biases both an Android and iPhone app are developed. Due to iOS built-in security settings the amount of data that is technically accessible to non-Apple apps is limited. Therefore, the initial project scope for data collection is restricted to attributes of which is certain that they can be tracked on all devices after users' approval. Since the Android operating system is relatively open in the sense that more system information can be accessed, it has been assumed that features accessible on iOS can also be tracked on Android phones.

In Appendix A follows an extensive list of potential attributes for data collection mentioned in research papers and whether they are out of scope for the first and most basic version of the reality mining app.

1.2.4 Competitive Advantage

Open Data Kit (ODK) is an open source software tool developed by researchers for researchers with the purpose of survey data collection. Although its feature set includes a GPS tracking system, it only stores the location once for every survey submitted. Moreover, its main focus is on collecting self-reported data from participants rather than automatic data-collection (i.e. reality mining). The same holds for commercial substitutes designed for experience sampling momentary assessment such as MetricWire, SurveyCTO, SurveySignal, LifeData, ESM Capture, Qualtrics and HarvestYourData.

Furthermore, the app used to collect data for The Quantified Self project (Moments) has proven to be able to track all attributes mentioned in Table 2. However, the data-export is far from clean and more importantly the app is designed to reduce the total time spent on the phone rather than for the sole purpose of data collection. For example, it does not include functionality to configure the data collection process, invite users, ask for formal consent and automatically transfer gathered data from participants to researchers (Appendix B).

1.2.5 Customer Segments

Given the aforementioned scope the following types of researchers can potentially benefit from using the tool (Raento et al., 2009) [9]: sociologists who study diffusion/migration/traffic flows, economists who want to compare objective data on consumer behaviour to subjective reports, psychologists looking at the ecological validity of psychological constructs and educational and organisational researchers that investigate individuals practices at the micro level.

2 Financial Plan

2.1 Expenses

Due to the high adoption rate of smartphones an important cost reduction is that people can participate using their own device (Sonck and Fernee, 2013) [10]. Nevertheless, some people may not be able to afford a smartphone which can lead to sample biases. For that reason loan phones should be available to lend out. In the same way an instruction video can illustrate inexperienced users how the smartphone and app should be used for research aims.

Description	**Type of Cost**	**Cost (estimate)**	**Source**
App development and testing (iOS)	Fixed	€ 51,700	(Crew, n.d.) [3]
App development and testing (Android)	Fixed	€ 51,700	(Crew, n.d.) [3]
Web app development (for configuration by researchers)	Fixed	€ 10,000	(Hagen, L. n.d.) [6]
2-3 minute instruction film	Fixed	€ 8,500	(Fix, J., 2010) [5]
Loan phones (incl. prepaid cards)	Fixed	€ 200/phone + € 10/month (cellular data)	(Raento et al., 2009) [9]
Server bandwidth and cloud storage	Variable	€ 84.70/month	(Rackspace, n.d.) [8]

Table 1: Estimated costs for product-development

What remains are mainly fixed development costs which drive up the expenses a lot (Table 1). Before investing large sums of money in software, however, it is highly recommended to start with the bare minimum also known as Minimum Viable Product (MVP) to test product hypotheses first. For example, a good starting point would be to initially use the existing Moments app, extract the raw JSON-file after the trial manually and repeat the data cleaning steps one by one.
Alternatively, to reduce costs a potential partnership could be entered into with Kevin Holesh, the developer of the Moments app.

2.2 Cost Benefit Analysis

In their paper Boase and Ling state that they found 41 journal papers that included the keyword "mobile phone" and were based on empirical research published between 2003 and 2010. That comes down to approximately 5 papers each year. The study carried out by Sonck and Fernee took place between March 2011 and March 2012 and included 150 participants.

The company SurveySignal developed by researchers from the University of Chicago Booth School of Business has a straightforward business model: a fixed rate of 10 dollar cents per SMS sent. In the pricing example on their website they assume 5 text messages are sent each day what implies that researchers are probably willing to spend \$0.50 (€0.42) per participant per day. For this analysis we suppose to have a similar business model: a fixed rate per participant per day of €0.42.

Combining these facts gives a total estimated revenue of €0.42 $\times$ 150 participants $\times$ 365 days = €22,995 for a single 1-year research study. It should be noted though, that reality mining research allows for large-scale and long-term studies as pointed out by Sonck and Fernee and Raento et al. Thus, the actual total number of participants may be higher and the research duration may be extended to over a year, so that the true revenue per study might be higher.

All in all, assuming all mobile phone experience sampling research studies utilise our reality mining tool the estimated sunk costs due to app and web development ($\sim$ €115,000) can be earned back in less than 1 year and thus the company could be profitable from year 2 onwards. Note that although a customer base of 5 studies in the first year may seem infeasible at first, the number of reality mining studies has grown significantly in recent years according to Web of Science statistics (Figure 1) which can compensate for a lower actual adoption rate.

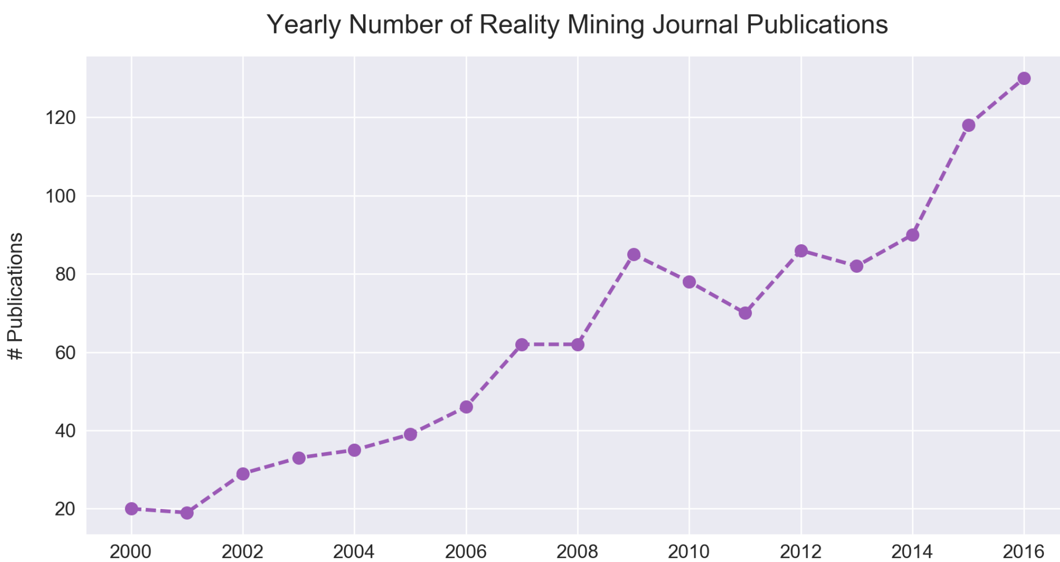

Figure 1: Upwards trend of number of reality mining articles published (2000-2016)

References

[1] Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). *Beyond self-report: Tools to compare estimated and real-world smartphone use.* PLoS ONE, 10(10), 19. https://doi.org/10.1371/journal.pone.0139004

[2] Boase, J., & Ling, R. (2013). Measuring Mobile Phone Use: Self-Report Versus Log Data. Journal of Computer-Mediated Communication, 18(4), 508519. https://doi.org/10.1111/jcc4.12021

[3] Crew (n.d.). How Much To Make An App. Retrieved from http://howmuchtomakeanapp.com/

[4] Doherty, S. T., Lemieux, C. J., & Canally, C. (2014). Tracking human activity and well-being in natural environments using wearable sensors and experience sampling. Social Science & Medicine, 106, 8392. https://doi.org/10.1016/j.socscimed.2014.01.048

[5] Fix, J. (2010). What does a corporate web video cost? 25 Factors (with prices) that affect corporate video production costs. Retrieved from: https://onemarketmedia.com/2010/03/03/what-does-a-web-video-cost-25-factors-with-prices-that-affect-video-production-costs/

[6] Hagen, L. (n.d.) How much it costs to build a web application. Retrieved from: http://djangostars.com/blog/how-much-it-costs-to-build-a-web-application/

[7] Hofmann, W., & Patel, P. V. (2015). SurveySignal. Social Science Computer Review, 33(2), 235253. https://doi.org/10.1177/0894439314525117

[8] Rackspace (n.d.). Pricing. Retrieved from: https://www.rackspace.com/cloud/servers/pricing

[9] Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones. Sociological Methods & Research, 37(3), 426454. https://doi.org/10.1177/0049124108330005

[10] Sonck, N.; Fernee, H. (2013). Using smartphones in survey research. The Hague: The Netherlands Institute for Social Research. Retrieved from

https://www.scp.nl/english/Publications/Publications_by_year/Publications_2013/Using_smartphones_in_survey_research_a_multifunctional_tool

A Appendix - Project Scope

Attribute	**In scope**	**Out of scope**
Number of calls		x
Length of calls	x	
Recording of calls		x
Recording of background noise		x
Number of SMS messages sent		x
Number of SMS messages received		x
Time on phone	x	
Number of phone pickups	x	
Location	x	
User interaction (e.g. pop-up to fill out a survey)		x
Intelligent context dependent triggers (incl. reminders)		x
Other devices in physical proximity		x
Search history on internet		x
Calendar events		x
Object detection with the phone's camera		x
Battery level	x	
Network coverage		x
Alarm status		x
Usage of apps	x	

Table 2: Overview of data attributes included in V1 mobile application

Note that although the collection of qualitative feedback for experience sampling through, for example, questionnaires is out of scope researchers can of course combine existing solutions with our product.

B Appendix - Process Diagram

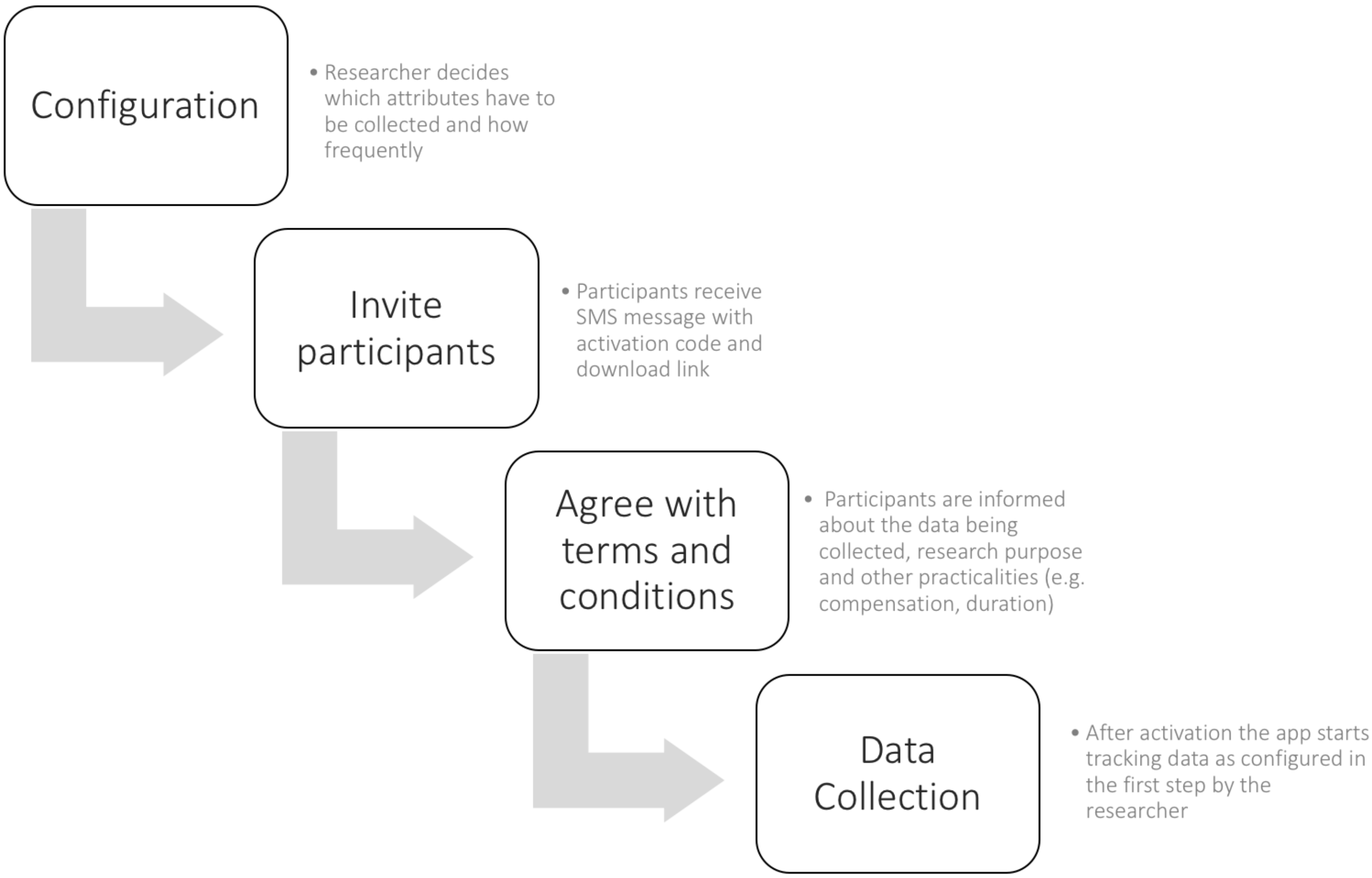

Figure 2: Research process: from configuration to data collection