Synthesis & Evaluation Of A Mobile Notification Dataset

Kieran Fraser
ADAPT Centre
Trinity College Dublin
kfraser@scss.tcd.ie

Bilal Yousuf
ADAPT Centre
Trinity College Dublin
yousufbi@scss.tcd.ie

Owen Conlan
ADAPT Centre
Trinity College Dublin
owen.conlan@scss.tcd.ie

ABSTRACT

Open-source mobile notification datasets are a rarity in the research community. Due to the sensitive nature of mobile notifications it is difficult to find a dataset which captures their features in such a way that their inherently personal information is kept private. For this reason, the majority of research in the domain of Notification Management requires ad-hoc software to be developed for capturing the data necessary to test hypotheses, train algorithms and evaluate proposed systems. As an alternative, this paper discusses the process, advantages and limitations with harnessing a large-scale mobile usage dataset for deriving a synthetic mobile notification dataset used in testing and improving an intelligent Notification Management System (NMS).

CCS CONCEPTS

•**Information systems** → *Open source software; Expert systems;*

KEYWORDS

Synthetic data creation; Notification management; Shared data; Open-source mobile notification dataset

1 INTRODUCTION

Gathering information through mobile devices is a trend that continues to grow in the research community. Due to a myriad of sensors available to researchers through mobile devices, and the rare rate of separation from their owners, they are an invaluable resource of rich user data [10, 20]. One such research application for datasets harvested in this manner is the training of intelligent systems for management of mobile notifications [8]. However, obtaining rich datasets, such as those containing mobile notifications and contextual user data, from mobile devices does pose a number of problems as there are a number of factors which impede the ease of acquisition. Some such factors include development, ethics, security, transparency and scale.

Many researchers in the domain of intelligent notification management have to invest time in developing ad-hoc mobile applications for gathering mobile notification data due to the limited number of open source datasets available to them which contain the unique data points required to test their hypothesis [11, 16, 19, 22]. For example, there are few open source notification datasets which have sufficient granularity for context driven delivery - those which contain both notification and user data. This has led to a number of researchers investing time in building applications which capture the necessary data for their evaluations. While the time invested in building and capturing the data is worthwhile, the reuse of such tools and datasets is restricted due to ethical constraints placed on them during collection. Users' give permission for their data to be harvested and used for very specific experiments and evaluations which may not cover other similar research efforts. Hence, a completely new dataset may be required if the parameters of experimentation change or data is to be used differently. While this transparency is necessary as it ensures the ethical use of users' collected data - it is also time consuming to implement and can be a big barrier to research efforts. Scaling a mobile application which captures notification and contextual user data is a slow process and getting a large number of users' to participate and use an application which is harvesting their personal data is difficult, although rewarding if achieved [6]. For this reason there is a lack of large-scale datasets freely available. This in turn hampers progress on building effective notification management systems (NMS) which can be thoroughly trained, evaluated and tested.

A potential alternative to using notification data captured through ad-hoc mobile applications is to generate a synthetic notification dataset. A synthetic dataset can be shared and reused as the same ethical implications with using real user data doesn't apply. It can also be modified and scaled according to the needs of the research. While the flexibility and freedom of use are advantageous, some disadvantages also materialize with synthetic data. It may not, for example, accurately describe real world scenarios resulting in the data not being fit for the purposes of training or evaluating an intelligent system. This paper discusses the creation of a synthetic dataset which is derived from a real-world mobile usage study for the purposes of improving a mobile notification management system as well as other advantages, challenges and potential opportunities synthetic dataset creation offers. As the data is grounded in authentic real world data, it strikes a compromise between accuracy and flexibility which enables it to improve an intelligent system.

2 RELATED WORK

Similar work has also been done to improve a Notification Management System through the use of synthetic data. Corno et al. [4] combine synthetic with authentic data to create a hybrid dataset with which to train their NMS using a variety of machine learning algorithms. The benefit in this case was that training and testing could be carried out quickly and efficiently in order to verify the potential of the machine learning algorithms when applied in the context of notification management. The focus could remain on the development of the intelligence of the system as opposed to building an application to gather data.

DOI:

Another paper evaluates the use of synthetic data for training and testing of fraud detection systems (FDS). Lundin et al. [15] discuss the difficulty in capturing the large amounts of data necessary to build an intelligent FDS. It is noted that most studies don't put enough focus on their test data and that the majority of the data used in experimentation, training and evaluation is authentic but manipulated and fit-for-purpose making it unfit for other research studies. The proposed solution is a synthetic dataset although it is conceded that there is a drawback with this synthetic datasets as the possibility exists that the it will not draw parallels with the real world. To counter this, the synthetic dataset is derived from a small authentic dataset and then scaled. Using the authentic data as a seed for the synthetic dataset, it enables the data to have an element of real world ground truth while still enabling the researchers to add properties and scale the amount of data to fit their needs.

Comparably, Whiting et al. also create a synthetic dataset for testing and evaluating information analytic tools. This dataset is created via a known ground truth which then spawns data cues which are inserted into the dataset with a certain expressivity. The number of cues and subtlety of their representation can also be controlled giving creators a level of flexibility over the test dataset.

In contrast with those using synthetic datasets, a number of studies in the domain of mobile notification management opted to build ad-hoc applications which capture data *in-the-wild* [7, 16, 18, 21]. The limitation with this is the data is collected for a specific purpose which users' agree to from the outset. If the parameters of the experiments are to change, the data can't adapt to fit the new purpose. Scaling this data is also difficult as it is totally dependent on the engagement of the user base. Alternatively, synthetic datasets potentially only need a modest sized real-world dataset as a seed which can then be scaled to suit the needs of the experiment. A balance needs to be struck between size and quality if the synthetic dataset is to perform however.

3 EXISTING DATA

At the time of writing, there are few, if any open-source mobile notification datasets in existence. The difficulty being that notifications are inherently personal and generally contain sensitive information which is unethical to record, use and share. Research in the domain of notification management has therefore contained, thus far, many ad-hoc mobile applications which capture a limited number of data points relating to user's incoming notifications and proceeds to abstract this data to preserve its sensitive nature [8, 17, 18]. The raw information contained within the notifications is discarded and intelligent systems are trained on abstracted data which tends to highlight the emphasis currently being placed on ensuring the utmost user privacy and protection.

Alternatively, there are a number of mobile usage and behavioral datasets which have been made available to varying degrees. These datasets capture a wide ranging number of data points harvested from mobile sensors, user interactions and surveys pertaining to mobile usage. While these datasets don't explicitly capture notification data they do capture general mobile interactions such as SMS, calls and application usage. This data therefore has potential use in creating a synthetic notification dataset. The following is a brief description of the more popular datasets:

(1) **Nodobo Dataset** [2] - this was a mobile usage dataset gathered in 2010/11 by the University of Strathclyde. The features included in the dataset are calls (13035 records), messages (83542 records), presence (5292103 records), wifi and cell towers - all data is anonymized. It's freely available for use with the only stipulation being that researchers cite the relevant Nodobo publication. The dataset is downloaded in the form of multiple CSV files.
(2) **Worldwide Mobile App User Behavior Dataset** [14] - this was a worldwide survey of mobile application user behavior encompassing 15 countries and 10,000 participants in 2012. Survey questions include those surrounding demographics, age, education level, personality (Big-Five), app usage, triggers for downloading apps, why they abandon apps and so on. This dataset is also freely available with the stipulation of citing the relevant publication if used. The data is downloaded in the form of an Excel file.
(3) **Device Analyzer Dataset** [23]- this is a growing dataset harvested via an Android application which logs statistics of phone usage in the background. It was developed in the University of Cambridge and has 31,031 contributors with over 100 billion records of smartphone usage from over 17,000 devices around the world. The full list of data points collected can be found in the reference, though it contains various mobile events such as time phone is charging, time pictures taken, sensor readings and so on. The dataset is available for free for academic researchers and is obtained via email once a brief description of use is given.
(4) **CrowdSignals.io Dataset** [5]- this is a community which aims to generate ongoing labeled mobile and sensor data. The data is free for research purposes 12-18 months after collection, before which it is accessible for a fee. The datasets include features involving geo-location, communications, user interactions, social connections, user-provided ground-truth labels and survey feedback. At this point there are only sample datasets available as the data collection period is ongoing.
(5) **Mobile Data Challenge (MDC) Dataset** [12]- this was a data collection campaign run by Idiap and NRC-Lausanne from 2009 to 2011 whereby 200 participants had their mobile device behavior and social network detail collected. The dataset is restricted to non-profit organizations and requires a process of establishing a site manager who acts as a legal representative for an eligible institution before the data is passed on. The data is also considered to be only accessible to permitted staff within the eligible institution.
(6) **Friends & Family (F&F) Dataset** [1] - this was an experiment studying how people make decisions with certain social elements involved and how the decisions can be improved. The dataset includes over 130 participants, was run in 2010/11 and includes features such as call/SMS logs, applications installed, relationships, activities, events, mood, personality (Big-Five) and so on. The data was collected *in-the-wild* via an *Android* app by the MIT Human Dynamics Lab and is now freely available with the stipulation that relevant publications are cited, the data isn't distributed

to others and re-identification of the participants is not attempted. The dataset is downloaded as a set of CSV files.

All the datasets discussed have varying degrees of quality, and none adhere to a standard set of features or collection methods. The form in which they are distributed is also diverse, although separating the data into multiple CSV files appears to be the most common method. A recent analysis on mobile datasets delves further into the application of these datasets in research along with other mobile datasets which are not made freely available - it also highlights the importance of such large-scale datasets in the research community and their value across multiple domains [3].

4 SYNTHETIC DERIVATION FROM F&F DATASET

As previously discussed, the availability of open-source notification datasets is low. Effective training and evaluation of a NMS is difficult to achieve without a significant amount of data to test with. Creation of a mobile application to gather the necessary data is inefficient as application development and an active user base is difficult to scale quickly. Aside from this, ethical implications are also a deterrent in capturing notifications *in-the-wild* as the necessary data to gather is highly personal to the user. The notification content could also potentially implicate people not directly involved with the data gathering process - for example, a notification sent from a user's friend could be highly sensitive in nature such that they would not want the information shared.

As an alternative to capturing notification data *in-the-wild* a synthetic notification dataset is proposed for use in improving a NMS [8]. In order for the dataset to be useful, it must draw parallels with real-world data. For this purpose, the synthetic notification dataset is derived from a real-world mobile usage study - the F&F dataset discussed in the previous section [1]. The F&F dataset was chosen due to its inclusion of Call and SMS logs, which can be interpreted as incoming notifications, as well as application usage and a rich store of contextual user data such as calendar events, moods, personality types and relationship perceptions. No other dataset gives such a granular and holistic view of the user both with reference to their device and without.

The derivation process can be split into the following stages: feature analysis, cleaning, assumptions and scaling.

4.1 Feature Analysis

The F&F dataset contains a number of features split by category: Call log, SMS log, Location, Apps in phone, Running Apps, Acceleration, Battery status, Monthly surveys, Weekly Surveys, Social network structure and Funfit surveys. For example, the features in the call log category are: participant id (which remains fixed throughout the categories allowing them to be merged), date and time of the call, participant id of the calling device and an anonymized MAC address. The first step in deriving the synthetic notification dataset was to identify the categories and features in the F&F dataset which were applicable. As the dataset being created was to be used in improving an existing NMS, NAbs, the applicability of features for the synthetic dataset was dependent on the features used when previously training and evaluating this system. Table 1 illustrates the F&F categories chosen for creating the synthetic dataset and the NAbs feature they are applicable for seeding.

Table 1: Friends & Family categories used for inferring NMS dataset features.

Friends & Family Category	NMS dataset feature
Accelerometer readings	Not used
Apps installed	Notification app
Apps running	Not used
Bluetooth proximity	Not used
Battery Information	Not used
Call Log	Notification time/date
GPS	Not used
WiFi access points nearby	Not used
SMS Log	Notification time/date
Big 5 personality answers	Not used
Relationships (Couples & Friendship)	Sender ranking
Survey monthly, weekly and daily	Subject, Sender, App, App ranking, User events/ activities

4.2 Cleaning

The downloaded form of the F&F dataset is comprised of a number of CSV files (approximately one per category). As some of the categories consist of qualitative survey data, the form in which it is presented isn't actionable. Therefore, a number of data manipulations were carried out to join the data together before it was assembled into a usable model. To achieve this the CSV files were imported as individual tables into a SQL database. A number of queries were then run to extract the applicable data and manipulate it into a form which could be easily extracted for the user model of NAbs.

4.3 Assumptions

The F&F dataset does not contain all the explicit features required for the NAbs NMS, therefore assumptions are placed on existing features to transform them into applicable NAbs features and thus form a synthetic notification dataset. The necessary features are:

(1) Notification data - Sender, subject, application, date/time.
(2) Activity data - Events/activities the user is engaged with at the time of notification delivery.
(3) User preference data - A general importance a user gives to possible terms of the notification features.

The following is a brief description of the inferences made for each feature of the synthetic dataset derived from the F&F dataset:

Notification date/time - This feature was inferred using the SMS and Call log categories, each of which had incoming date and time features. By using this frequency of incoming communication alerts, the synthesized data remains grounded in real-world scenarios.

Notification Sender & Ranking - These features refer to the sender of the notification and the relationship importance value the user gives to that sender. The SMS and Call log give the sender id value. By joining the sender id with corresponding features from the Relationships category, the sender can be identified as a partner or a friend. The Relationship category is made up of two sub-categories: the Friendship category is comprised of relationship ratings given by each user to other users. The Couples category is comprised of a mapping of one user id to another to signify both users are a couple. Assumptions made for populating the sender and corresponding sender ranking features are:

(1) If a sender cannot be found within the Friendship or Couple datasets, the sender is a stranger and is assigned a user importance rating of 0.
(2) If a sender is found in the Friendship dataset and has a rating greater than 0, the sender is classified randomly as either a colleague or a friend. This was done simply to add context to the notification and aid in inferring the subject (also discussed in this section).
(3) If a sender is found in the Couples dataset then the sender is classified as family and the user rating is 10. This user rating is assumed and may not always reflect a real-world situation.

Notification Subject & Ranking - These features refer to the subject of the notification and it's corresponding importance to the user. No information on the subject was recorded in the SMS or call categories, therefore the subject had to be inferred by other means. A fixed set of subjects was therefore assumed and the sender of the notification was used to infer which subject of the set was chosen. The subject set is comprised of {family, work, social, interest}. This set of subjects was not chosen randomly but based on an evaluation of the *in-the-wild* notification datasets previously used with the NAbs NMS and also based on the Survey categories of the F&F dataset which groups activities by these terms enabling the mapping of inferred subjects to ground-truth data. The assumptions are as follows:

(1) If the sender of the notification is a family member, the subject is family.
(2) If the sender of the notification is a colleague, the subject is work.
(3) If the sender of the notification is a friend, the subject is social.
(4) If the sender of the notification is a stranger, the subject is interest.

Limitations do occur when making assumptions such as these as a work colleague could in reality also send a social notification for example, hence this is not a complete representation of a real world dataset. Additional nuanced assumptions could in future be added to enrich the data so that it covers additional scenarios however, for the purposes of this study the four assumptions mentioned above for inferring the subject are adequate. the subject ranking was not derived from the F&F dataset. It may have been possible to assign a ranking from each user to each subject term of work, social, interest and family by making assumptions based on their personality or frequency of incoming notification senders (e.g. many notifications from family members might signify a higher ranking), however it was decided that for the purpose of use with the NAbs NMS it was unnecessary as an ad-hoc subject ranking allows for control over the personality of the synthetic user, thus enabling exploration of scenarios.

Table 2: Mapping of app categories to notification subjects.

App Category	Subjects
Games	Family, Interest
Lifestyle	Family, Interest
Shopping	Family, Interest
Communication	Family, Interest, Social, Work
Entertainment	Interest, Social
Phone Personalization	Interest
Productivity	Work
Social	Social
Other	Interst

Table 3: Mapping of event types to assumed relevant times.

Event Type	Time
Most look forward to	5pm - 8pm
Most enjoyed	5pm - 7pm
Cinema	8pm-9pm
Movie	9pm-10pm
TV	5pm-8pm
Restaurant	7pm-9pm

Notification Application & Ranking - These features refer to the app the associated notification was delivered through and the importance value placed on it by the user. For the synthetic dataset there is no explicit app for each notification as the notifications in the synthetic dataset were originally SMS messages and calls therefore the app for a particular notification is chosen based on the inferred subject of the notification (which was chosen based on the sender). A mapping of subject to app category was assumed and used to infer which app category was chosen for a given notification (Table 2). Once the app category was chosen, a user's known apps were retrieved from the Running Apps and Survey categories in the F&F dataset. Apps matching the chosen category were added to a pool from which one app was drawn randomly to be the app for the given notification. The user ranking of importance was inferred using the Survey dataset which asked users to rank their most used and favorite apps.

4.4 Scaling

The last set of features for the synthetic dataset relate to the receiver of the notification and their contextual situation when the notification was received. The qualitative surveys in F&F enable the creation of calendar "events" and "activities" due to the granularity of data provided on user behavior. Other datasets described Section 3 do not contain contextual user data as detailed as this, hence the reason for choosing F&F. Even with this data, users' calendars are sparse hence an additional step is to scale the number of events a user has depending on other attributes found in the F&F dataset.

User Events/Activities - These features describe events and activities a user has planned or is currently doing. They include the start and end time of the event/activity, who is with them during it and the subject it pertains to. There is no explicit information regarding a userfis current activity at the time of all incoming calls and SMS messages in the F&F dataset hence, a number of assumptions have to be made regarding usersfi activities. The Survey categories in the F&F dataset gather information regarding some events and activities that a user partakes in during the weeks of the study. For example, it records visits to the cinema, restaurants and nights in front of the TV. It also records which groups of people the activity or event is associated with. For example, userfis specify if the event is attended with family members, colleagues, friends or alone. These are used to create a sparse calendar for users. Additionally, the Survey category contains information regarding whether or not the user is a student. Based on this a number of assumed fixed events are added to a users calendar to scale the amount of contextual information available. The event assumptions are as follows:

(1) If the user is a student - they have college lectures from 9am to 5pm each weekday with a break of 1 hour for lunch at 12pm.
(2) If the user is not a student they are assumed to have work from 9am to 5pm each weekday with a break of 1 hour for lunch at 12pm.
(3) For events taken from the Survey category of the F&F dataset, the day of the week is chosen randomly and the time and duration are chosen based on the type of event (Table 3).

The last manipulation to be done on the synthetic dataset is to ensure the event dates align with the incoming notifications. In some cases the events start after the last known incoming notifications (call or SMS message) took place. As the events are simply a sample of the habits of the particular user, and it is the frequency and subject they represent which is important, they are simply shifted in time to align with the start of the incoming notifications.

The completed synthesized dataset is made up of 32 users and 11395 notifications and 3148 calendar events. Additional changes can be made to the dataset depending on the needs of the research. For instance, the categories used for the notification attributes (e.g. *family* or *friend* for sender, *work* or *social* for subject) are open to interpretation and can be chosen based on the use of the dataset. The F&F dataset is simply a seed of real world data which is used to spawn a large scale synthetic dataset. There are a number of limitations which put the dataset at a disadvantage to those which capture notification data *in-the-wild*. For example, the NAbs NMS would require some feedback from the user on the preferred delivery time of a incoming notification in order to determine the whether or not it correctly inferred an appropriate delivery time. This is not possible to synthesize at scale as it is not fully understood what feature values would govern such feedback. The dataset can still be applied to test and derive insights from the performance of intelligent systems however as briefly discussed in the following section.

5 CASE STUDY: NABS

The synthetic notification dataset, the creation of which is outlined in the previous section, was evaluated [9] using an existing Notification Management System (NMS), NAbs [8]. This NMS infers a preferred delivery time for a notification based on the *sender*, *subject* and *app* attributes as well as the user's calendar of scheduled events.

A clear limitation of the synthetic dataset is the missing feature *preferred delivery time*. This is the feature that NAbs attempts to infer, but without the explicit value provided by the user for each incoming notification, it is difficult to ascertain whether or not the correct inference was made. There is no feature in the F&F dataset from which the *preferred delivery time* could potentially be derived. If the dataset had included missed calls, or a time when a SMS message was read and not just delivered then the feature could possibly have been derived. An alternative method was therefore proposed for using the synthetic dataset to improve the NAbs system. The NAbs system operates using a Mamdani Fuzzy Inference System (FIS). This means that the inferred delivery time is calculated using a combination of fuzzy membership functions and a fuzzy knowledge base. This fuzzy knowledge base is made up of a number of semantic rules such as:

IF sender IS *important* AND subject is *very important* THEN *deliver now*

Usually these rules are fixed manually in the FIS by a domain expert. However, in the domain of notification management, the only person deemed expert to derive rules of when a notification should be delivered is the user themselves. However, asking a user to compose a knowledge base full of rules which span many scenarios is unrealistic - hence some form of machine driven personalisation must take place to create and maintain the knowledge base. In the previous evaluation of NAbs [8], the knowledge base contained a small number of general rules. One of the limitations of the work was cited as a lack of dynamism in the knowledge base to reflect a user's changing needs. For example, during the hours of 9 to 5 a user may place greater importance on incoming work notifications while outside of these hours, social notifications would take precedence. There are of course a number of situations which would contradict this assumption and it would also vary greatly on the person involved. However, for the purposes of improving the generic knowledge base, this assumption is adequate - it can also be seen as a means of addressing the cold start problem if no former data is known about the user's preferences. Hence, the synthetic dataset is used by NAbs to improve the FIS's knowledge base by learning which rules most satisfy a generic assumption regarding which notifications a user has precedence for at a particular time of day.

The generic assumption, *work-play-split*, is:

Users prefer to receive most work-related notifications straight away if they arrive within working hours and prefer to receive most social-related notifications later during these same hours.

By assuming each user in the synthetic dataset falls under this stereotype, NAbs can be trained over a diverse set of scenarios to converge to *preferred delivery times* recommendations which match the *work-play-split*.

NAbs was trained with the synthetic dataset using a population based search algorithm named Particle Swarm Optimisation (PSO), which enabled a set of optimal rules be derived. These optimal rules were then applied to the fixed knowledge base in the FIS and the performance of NAbs, within the bounds of the *work-play-split*, was evaluated on a real-world notification dataset captured *in-the-wild*.

The results of the evaluation were that the derived rules obtained through use of the synthetic notification dataset do improve the performance of the NAbs NMS when the real-world data converges toward the assumption of the *work-play-split*. A second experiment was carried out to dynamically improve the fuzzy membership functions in the FIS of the NAbs NMS and the results were similar. Hence, the synthetic notification dataset added value toward understanding and improving an intelligent system.

6 REUSING THE SYNTHESIZED DATA SET FOR FUTURE WORK

This section describes how the synthetic derivation from the F&F data set discussed in this paper will be used by the ADAPT Centre in its future work. It also highlights that this data set will be made available to ths research community.

6.1 Usage within the ADAPT Centre

Within the ADAPT Centre, this data set will be used in a user trial enabling individuals to scrutinise the notifications through visual narratives and to visually control and update the semantic rules (presented in section 5) encoded by NAbs and inferences applied to the data. These inferences include the ranking of notification senders, subjects and applications discussed in section 4.3.

Research has shown that visual narratives support users in understanding data presented [13, 24, 25]. The visual narratives will be used to guide individuals through the notifications supporting user analysis and exploration. This will include analysing the inferences applied and rules encoded by NABS to determine when to deliver the notifications.

The user trial will also support users to visually control and update the inferences and rules, thereby having direct user involvement in the personalisation of notification delivery.

6.2 Availability of the Synthesized data for the Research Community

The synthesized data set will be made available to the research community to access, use and update through the ADAPT Centre, however the data assumptions that have been applied to the data during the synthesis process, and which described in sections 4.3 and 4.4 above need to be noted.

ACKNOWLEDGMENTS

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

REFERENCES

[1] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. Social fMRI: Investigating and Shaping Social Mechanisms in the Real World. *Pervasive Mob. Comput.* 7, 6 (Dec. 2011), 643–659. DOI: https://doi.org/10.1016/j.pmcj.2011.09.004

[2] Stephen Bell, Alisdair McDiarmid, and James Irvine. 2011. Nodobo: Mobile phone as a software sensor for social network research. In *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*. IEEE, 1–5.

[3] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. 2015. A survey of results on mobile phone datasets analysis. *EPJ Data Science* 4, 1 (2015), 10.

[4] Fulvio Corno, Luigi De Russis, and Teodoro Montanaro. 2015. A context and user aware smart notification system. In *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*. IEEE, 645–651.

[5] CrowdSignals.io. 2017. A massive new mobile data collection. http://crowdsignals.io/. (2017). Online; accessed 19 April 2017.

[6] Denzil Ferreira, Vassilis Kostakos, and Anind K Dey. 2012. Lessons learned from large-scale user studies: Using android market as a source of data. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 4, 3 (2012), 28–43.

[7] Joel E. Fischer, Chris Greenhalgh, and Steve Benford. 2011. Investigating Episodes of Mobile Phone Activity As Indicators of Opportune Moments to Deliver Notifications. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. ACM, New York, NY, USA, 181–190. DOI: https://doi.org/10.1145/2037373.2037402

[8] Kieran Fraser, Bilal Yousuf, and Owen Conlan. 2016. A Context-aware, Info-bead and Fuzzy Inference Approach to Notification Management. In *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference sponsored by IEEE (Published by IEEE Xplore) (IEEE UEMCON 2016)*. New York, USA.

[9] Kieran Fraser, Bilal Yousuf, and Owen Conlan. forthcoming. An In-the-wild & Synthetic Mobile Notification Dataset Evaluation.

[10] Gabriella M Harari, Nicholas D Lane, Rui Wang, Benjamin S Crosier, Andrew T Campbell, and Samuel D Gosling. 2016. Using smartphones to collect behavioral data in psychological Science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science* 11, 6 (2016), 838–854.

[11] Nicky Kern and Bernt Schiele. Context-aware notification for wearable computing. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWCfi03)*. 223–230.

[12] Niko Kiukkonen, Jan Blom, Olivier Dousse, Daniel Gatica-Perez, and Juha Laurila. Towards rich mobile phone datasets: Lausanne data collection campaign. (????).

[13] Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. 2013. SketchStory: Telling more engaging stories with data through freeform sketching. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2416–2425.

[14] Soo Ling Lim, Peter J Bentley, Natalie Kanakam, Fuyuki Ishikawa, and Shinichi Honiden. 2015. Investigating country differences in mobile app user behavior and challenges for software engineering. *IEEE Transactions on Software Engineering* 41, 1 (2015), 40–64.

[15] Emilie Lundin, Håkan Kvarnström, and Erland Jonsson. 2002. A synthetic fraud data generation methodology. *Information and Communications Security* (2002), 265–277.

[16] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing Content-driven Intelligent Notification Mechanisms for Mobile Applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 813–824. DOI: https://doi.org/10.1145/2750858.2807544

[17] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing content-driven intelligent notification mechanisms for mobile applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 813–824.

[18] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding Peoplefis Receptivity to Mobile Notifications.

[19] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Reducing Users' Perceived Mental Effort Due to Interruptive Notifications in Multi-device Mobile Environments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 475–486. DOI: https://doi.org/10.1145/2750858.2807517

[20] Earl Oliver. 2010. The challenges in large-scale smartphone user studies. In *Proceedings of the 2nd ACM International Workshop on Hot Topics in Planet-scale Measurement*. ACM, 5.

[21] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 897–908. DOI: https://doi.org/10.1145/2632048.2632062

[22] Yujue Qin, Tanusri Bhattacharya, Lars Kulik, and James Bailey. 2014. A Context-aware Do-not-disturb Service for Mobile Devices. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14)*. ACM, New York, NY, USA, 236–239. DOI: https://doi.org/10.1145/2677972.2678003

[23] Daniel T Wagner, Andrew Rice, and Alastair R Beresford. 2014. Device Analyzer: Large-scale mobile data collection. *ACM SIGMETRICS Performance Evaluation Review* 41, 4 (2014), 53–56.

[24] Bilal Yousuf and Owen Conlan. 2014. Enhancing Learner Engagement through Personalized Visual Narratives. In *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on*. IEEE, 89–93.

[25] Bilal Yousuf and Owen Conlan. 2015. VisEN: Motivating Learner Engagement Through Explorable Visual Narratives. In *Design for Teaching and Learning in a Networked World*. Springer, 367–380.