

Discovering Types of Smartphone Usage Sessions from User-App Interactions

Björn Friedrichs
School of Computer Science and
Informatics
Cardiff University
Cardiff, UK
FriedrichsB@cardiff.ac.uk

Liam D. Turner
School of Computer Science and
Informatics
Cardiff University
Cardiff, UK
TurnerL9@cardiff.ac.uk

Stuart M. Allen
School of Computer Science and
Informatics
Cardiff University
Cardiff, UK
AllenSM@cardiff.ac.uk

Abstract—Understanding how and why people use their smartphones has enabled use cases ranging from correlating behaviour with psychological states through to on-device tasks such as app recommendations. However, being able to effectively and pervasively capture usage behaviour is challenging due to the wide range of functions, apps and interactions that are possible. In this paper, we examine how embedding physical user-app activity (e.g., taps and scrolls) can provide a rich basis for summarising device usage. Using a large dataset of 82,758,449 interaction events from 86 users over an 8-week period we combine feature embedding and unsupervised learning to extract prominent interactions within clusters of smartphone usage sessions. We find that high-level features such as session length, unlock state, and app switches are not representative of these clusters and can give a false sense of similarity or dissimilarity between sessions. The results motivate further exploration of the utility of using user-app interaction behaviour as the basis for the aforementioned use cases.

Index Terms—smartphone, smartphone usage, mobile, user behaviour, interaction

I. INTRODUCTION

Smartphone usage has been of rising research interest as a means of finding linkages between behaviour and cognitive states such as focus, mood, and interruptibility, as well as in designing intelligent systems surrounding future usage (e.g., next app prediction [1]). Previous works have sought to reduce and summarise usage data to determine common ‘types’ of sessions (e.g., [2, 3]) and have found that while usage is diverse, distinct but variable clusters can be found. Similar findings have emerged considering ‘types’ of smartphone [4] or web [5] users, and identifying personal habits in order to produce content recommendations [6]. However, this has primarily been limited to using high-level information such as screen-time, app-time, and lock-state (e.g., [7, 8]), sometimes in combination with ambient information (e.g., [9]), which may not be representative of the physical activity and engagement taking place.

While high-level features result from series of low-level events, they are not a proxy for the intricacy of interaction that takes place during smartphone use. This can create the potential for sessions with comparable length or with equivalent apps used to be considered similar, despite displaying very different levels of interaction or engagement (see example in Figure 1), and vice versa. This creates the ability for usage

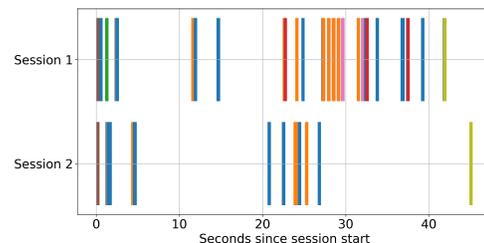


Fig. 1: Example of 2 sessions from the dataset of similar length with diverse frequency of interaction events (e.g., taps, scrolls) and idle time. Session 1: 53 events, 9 types, Session 2: 17 events, 5 types. (Different event types represented by colours)

to be misrepresented and motivates capturing and embedding additional interaction and engagement information. This is a challenge in its own right due to the highly variable and unlabelled task structure within individual applications. Previously insight has shown individuality of some aspects of usage (e.g., [8]), and homogeneity of other aspects within groups or across a population (e.g. app switching behaviours [10, 11] or notification management [12]).

A key task across these is being able to effectively summarise and capture a representation of usage over a time period. The data underpinning these works have primarily consisted of application driven information streams including screen time (e.g., [2]), the amount of apps used or how they are used in sequence/revisited [8]. However, further contextual factors have also been examined including battery level [9], Bluetooth use [9], cellular or WiFi signal status and strength [9], GPS location [13], received text messages [13, 2] and other notifications [12]. We note that previous works infrequently embed low level interaction events (e.g., taps, scrolls, and typing).

In this paper, we propose a new method for embedding smartphone usage behaviour that accounts for the diversity and noisiness in physical interaction events (e.g., taps, scrolls, and typing), but without the need for understanding the task structure of individual apps. Using a large dataset of real-world smartphone usage, we compare the method against various high-level features in discovering types of smartphone usage

TABLE I: An overview of prior studies that define what a session is and how it uses that information to determine forms of usage.

Study	Session boundaries	Frequency of data collection	Use cases
Kang et al., 2011 [9]	Start to end of network use or voice call	Periodical polling	Grouping into sessions and battery life prediction
Soikkeli et al., 2011 [3] Oulasvirta et al., 2012 [7]	Application launch to quit Unlock to lock event	Not specified Polling up to every 3 seconds	Sessions in a user context Differentiate between rapid, repeated and specifically triggered sessions
Ferreira et al., 2014 [10]	Application boundaries	Real-time application events	Identification and supplementation of micro usage
Jones et al., 2015 [8]	Application launch to quit and screen unlock to lock	Real-time application events	App revisitation analysis
Banovic et al., 2014 [13]	Screen on to off (merged if less than 5 seconds)	Screen and application events, notifications, GPS and activity data	Glance, review and engage categorisation
Jesdabodi et al., 2015 [14]	Active device use bounded by mean pause time	Mix of Real-time application events and polling.	Extracting usage states (e.g. routines or long stretches)

sessions using unsupervised learning. In doing so we address the following research questions:

RQ1 Do high-level features surrounding screen-time, lock-state, and app switches misrepresent the complexity of smartphone usage?

RQ2 To what extent do features surrounding user-app interactions better capture distinct ‘types’ of usage sessions?

The results show that summarising usage sessions using lower-level interaction data in this way can create distinct clusters of sessions in comparison to higher-level time and app-based features. This suggests that these high-level features are only weakly representative and overgeneralise usage behaviour. Overall, this motivates exploring the utility of this representation for use cases where it is required to draw conclusions between usage and psychological states such as advancements in designing intelligent recommendation systems.

High-level features such as session length and unlock time are easy to obtain and therefore the first step towards examining user behaviour. However using a smartphone is a complex combination of low and high-level interactions. The literature is currently still limited in analysing more of the low-level interactions and how these can be effectively embedded into a representation that summarises behaviour. Towards this, there have been advances for specific tasks such as behavioural biometrics for authentication purposes [15] or to retrieve emotional states from keystrokes [16]. While these studies make use of low-level interactions to find answers to their respective questions they do not give insight in how they might be used for a more general overview of the respective user and session types.

II. DATA OVERVIEW

We use the Tymer dataset [17] as the basis for this study. It consists of app-window, device and user-interaction events collected from 86 users over an 8-week period. In total it contains 82,758,449 events of 19 types, including screen on / off, tap, typing and scroll events (see Table II).

Several steps were taken to clean and structure the data for analysis. Firstly, any duplicate events in the data were removed as were scroll events that occur within 200ms of a previous

scroll event (as they are generated verbosely and a user is unlikely to react within this time [18]). After this, we group all events that occur between pairs of screen on and screen off events (N=415,505) to define a ‘session’, based on the work by Banovic et al. [13]. As we are interested in capturing all defining characteristics of device interaction within sessions, we add additional pseudo-events representing moments where the user is ‘idle’. An idle event is created where there is a time gap of more than 30 seconds between any two other events. Inconsistent entries in the dataset (N=13,441) where screen events are followed by another of the same type (e.g., screen on followed by screen on) are removed. Sessions that started shortly after the previous session ended were not joined, but considered separately. Finally, sessions are removed from consideration if the time between pairs of screen on and off events (i.e. the session length) is longer than one hour (n=4), as these are considered extreme outliers that may have occurred due to gaps or faults in data collection.

We are interested in user interactions so remove the screen events themselves from further analysis. This highlights n=111,951 cases where the user turned the screen on and off without further interaction. These are on average only 5.97 (SD=5.86) seconds long which corresponds an automatic screen timeout of 5 seconds and could be qualified as either accidental or short glance sessions (e.g. for incoming notifications). These are arguably a ‘type’ of usage session in themselves but provide no basis for user interaction analysis so are excluded from the subsequent analysis, leaving N=290,109 sessions.

Frequency statistics across all sessions are shown in Table III with similar findings to previous studies (compare sessions from studies in Table I) in sessions being short (<3 minutes) and involving a small number of app switches. However, these distributions are long-tailed with a number of outliers and total event count in particular has notably high variance despite similarity in session length. This speculatively suggests that the lower-level events (as described in Table II) may play a substantial role in defining the characteristics of a session.

TABLE II: Event types and counts in the dataset.

Event categories	Event types, N=19	Description
Taps	Tap (23m), long tap (3m)	Direct touches on the screen.
Scroll	Scroll (30m)	Interface elements are scrolled off or into the currently visible screen region.
Typing	Text box change (11m), text selection (10m)	Typing and text-box interaction.
Other UI	View interaction (25m)	A list item being selected, or a check-box ticked, etc.
Window events	Window state change (3m)	Changes to application windows or the system UI.
Notifications	System notification (2m)	Notifications arrive or are updated
Screen	Screen on & off (removed), unlock (201k)	Changes to the device's screen state.
Power	Battery low/okay (2.8k/.4k), power on/off (.5k/.1k), device connected/disconnected (13k/12k)	Changes involving battery (e.g., low-battery) or power state (e.g., charging).
Idle	Idle (7,882,061)	Occurs every 30 seconds if no other events occur while the screen is on.

TABLE III: Frequency statistics of features. Session length and unlock time after screen on in seconds.

Features	Min	Max	M	Mdn	SD	Description
Session length	0.1	3595.8	148.3	37.3	342.7	Time between screen on and screen off.
Unlock time	0.1	3576.8	5.4	2	56.3	Whether the device was unlocked in the session and the time taken to do this.
App switches	0	1004	5.06	3	8.8	The number of switches between one app and another (e.g., via the home-screen).
Event count	1	104824	317.7	22	1320.3	The total number of user-app interaction, idle and device event.
Category count	0	16	1.9	2	1.5	The number of Google Play Store categories across the applications used.

III. REPRESENTING SMARTPHONE USAGE USING DEVICE ACTIVITY

We examine and compare different methods for summarising smartphone usage sessions at a high and low granularity. Firstly, we introduce a method for embedding device activity data (shown in Table II) into a summative vector-based representation for all sessions. During smartphone use some of the low-level event types are much more common than others, for example taps and scrolls would dominate any other event if we analysed them based on raw counts. Instead of frequency, we use the impact of particular event types to highlight relevant features in a session. To achieve this we utilise Term Frequency-Inverse Document Frequency (TF-IDF) [19], an embedding method first introduced for natural language processing tasks, that summarises term importance across a collection of documents. In the following section we detail how we adapted the method to session activity data.

A. Representing sessions as documents

The goal is to summarise sessions and the set of events within them into a single vector of features. The feature vectors are generated using TF-IDF where the inputs are *documents* (sessions) and *words* (the set of temporally sorted event types, N=19). This approach has previously been used to vectorise categorical data for usage sessions [20].

Formally, let e_1, \dots, e_{19} be the set of event types (Table II). We define the TF-IDF of event type e in session d as:

$$\begin{aligned} \text{Tf-idf}_{ed} &= \text{tf}_{ed} \times \text{idf}_e \\ \text{tf}_{ed} &= 1 + \log(\text{freq}(e, d)) \\ \text{idf}_e &= \log\left(\frac{1+n}{1+df_e}\right) \end{aligned}$$

where $\text{freq}(e, d)$ is the number of events of type e that occurred in session d , n is the total number of sessions and df_e is the number of sessions that contain an event

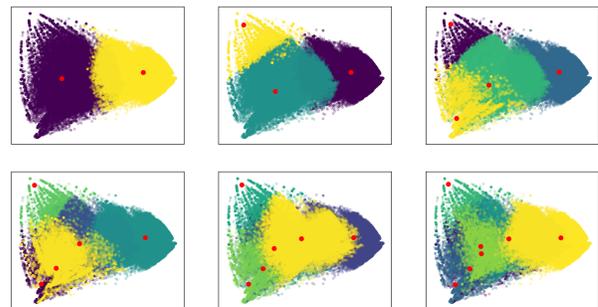


Fig. 2: Comparison of cluster independence across number of clusters ($k=2, \dots, 7$). Clusters are reduced to 2-dimensions using Principal Component Analysis

of type e . Each session d is then represented by a feature vector $f_d = (\text{Tf-idf}_{e_1d}, \dots, \text{Tf-idf}_{e_{19}d})$. The generated TF-IDF scores supply us with information about how impactful events are over others in terms of defining a usual usage session i.e. considering both the frequency and diversity of events.

B. Clustering sessions

Our goal is to identify groups of similar sessions based on their low level events by applying k-means clustering to their feature vectors. A rudimentary method to choose an appropriate number of clusters under k-means is to plot the inertia (sum of squared error per cluster) and choose the point of its 'elbow'. The inertia elbow in this case ended up relatively smooth but shows a small dip at $k=5$ hence we consider $k=5$ as the closest to best approximation but the elbow technique by itself is not completely conclusive.

To visualise our 19-dimensional clusters in 2d, we apply principal component analysis (PCA) (Figure 2), a technique to make multivariate data in large datasets more interpretable [21]. In this paper we apply PCA simply as a tool for visualisation, and therefore omit further analysis on its outcomes such as

TABLE IV: Frequency statistics of top 5 TF-IDF values.

Cluster 1 n=98267; 34.08%	TF-IDF M	TF-IDF SD	Count M	Count SD
Text Box	0.5	0.15	98.07	240.03
Text Selections	0.45	0.21	94.26	217.91
Scrolls	0.29	0.16	168.31	817.38
Taps	0.26	0.12	204.9	462.4
App Switches	0.17	0.11	0	0
Cluster 2 n=105571; 36.61%	TF-IDF M	TF-IDF SD	Count M	Count SD
Scrolls	0.31	0.29	78.75	452.7
Taps	0.23	0.21	8.03	37.22
App Switches	0.17	0.17	0	0
Unlocks	0.17	0.24	0.76	0.56
View Selections	0.16	0.28	58.14	412.82
Cluster 3 n=27115; 9.4%	TF-IDF M	TF-IDF SD	Count M	Count SD
App Switches	0.93	0.13	0	0
Taps	0.06	0.17	0.2	0.86
Unlocks	0.06	0.15	0.15	0.37
Notifications	0.02	0.11	0.04	0.21
App Switches	0.02	0.1	0	0
Cluster 4 n=29945; 10.38%	TF-IDF M	TF-IDF SD	Count M	Count SD
Idles	0.83	0.17	6.9	12.81
Unlocks	0.19	0.2	0.68	0.65
App Switches	0.07	0.14	0	0
Notifications	0.05	0.14	0.33	1.61
Taps	0.05	0.12	0.67	4.11
Cluster 5 n=27470; 9.53%	TF-IDF M	TF-IDF SD	Count M	Count SD
Notifications	0.81	0.19	16.3	101.7
App Switches	0.15	0.2	0	0
Taps	0.11	0.18	1.2	3.97
Scrolls	0.06	0.15	1.96	31.66
View Selections	0.05	0.17	6.03	67.38

TABLE V: Mean (SD) of high-level features per cluster. Session length in seconds.

	Session length	Event count	Switches	Categories
Cluster 1	229.6 (427.4)	748 (1997.6)	8.6 (11.3)	2.55 (1.4)
Cluster 2	114.7 (290.8)	161.8 (877.5)	4.8 (7.9)	2.06 (1.5)
Cluster 3	9.3 (10.5)	2.3 (3)	1.6 (1.7)	1.1 (0.3)
Cluster 4	227.7 (396.2)	10.4 (18.8)	0.9 (2.2)	0.5 (0.9)
Cluster 5	36.7 (127.9)	28 (163)	1.5 (2.8)	0.8 (1)

the variance. The choice of $k = 5$ as an appropriate value is also reinforced by using visual analysis of PCA clusters with clusters formed for $k = 2, 3, \dots, 7$. To further support this, we repeated the process using raw counts of each event type as the feature vectors. However, for $k = 5$, this resulted in 99.8% of all sessions being included in a single dominant cluster, demonstrating the importance of the TF-IDF approach.

C. ‘Types’ of usage sessions

We applied TD-IDF on the low-level events to all sessions and applied analysis based on higher level features to gain an understanding of the interaction between the different levels of

features. These values have an intrinsic value themselves, considering that every session is a combination of all interactions (events) by the user. Our first observation is that while users are very diverse in their overall usage, they are consistent in their types of session, with 98.84% of users having at least one session of each of the five types. This shows that while smartphone usage has previously been in the literature shown to be driven by individual patterns (e.g. [3]), we find that commonalities exist through these types of sessions.

Comparing the most important events to high-level features (Tables IV and V) identifies coherent patterns in each cluster:

- 1) Sessions whose main events are focused around text input and editing, also with a high session length, unlock time and multiple app switches (i.e. not just text messaging).
- 2) Comparatively long sessions with a focus on scrolls, taps and application switches, implying high activity for a prolonged time.
- 3) Very short sessions (median event count of one), with a heavy focus on an app switching, taps and notifications, consistent with glances after receiving a notification.
- 4) Sessions of 3-4 minutes, with low interaction/app switching, and large idle time between events, potentially being sessions that end up idling until the screen locks by itself.
- 5) Sessions with a strong focus on notifications but short on all high-level features could be indicative of a session that is changing media and triggering an internal notification.

Despite the process not including high-level features, the typical sessions within each cluster also have a defining set of typical high-level features. Kruskal-Wallis H-tests show that the distributions for each high-level feature do vary significantly across clusters, (session length: $H = 86111.4$, event count: $H = 147386.8$, app switches: $H = 86678.0$, category count: $H = 79491.7$, all $p < 0.01$), with Dunn post-hoc tests also showing significance between all pairs for all features except cluster 4 and 5 for event count.

IV. COMPARING CLUSTERS AGAINST SESSION FEATURES

The previous section detailed how low-level events can be formed into sessions and how those sessions can subsequently be clustered. Here we consider the overlap and divergence between low-level clusters and high-level metrics of sessions commonly used in the literature to summarise usage.

A. The role of lock-state

Another way to distinguish sessions is to consider the time it takes a user to unlock (the time from screen on to unlock event) or if a user unlocks their phone at all [8]. A fast unlock could be considered to be attached to a different kind of interaction, user or session than a slow one. In our records, approximately a third of all sessions do not have an unlock event attached to it (95,432 compared to 162,259 sessions). While this indicates a decently sized split, when mapping the sessions across the impact data of the TF-IDF results it seems that unlock time actually plays a much smaller, or even counterproductive role in grouping sessions. In the clusters the ratio of sessions having an unlock event are as follows in order: ~84.3%, ~71.7%, ~15%,

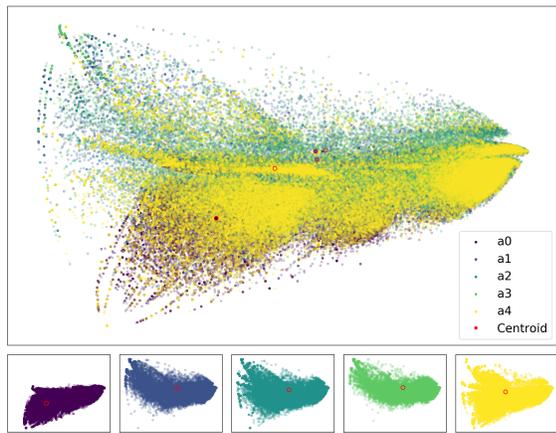


Fig. 3: All sessions clustered by only using their high-level features (session length, event count, unlock present, application switches and count of app categories) instead of TF-IDF values, plotted using PCA and showing the individual layers.

~62.8% and ~25.7%. In comparison, when sorting sessions by unlock time and then splitting it into 5 groups, 99.95% of all sessions and 100% of all sessions without an unlock end up in the first of those slices. Instead, with our method we can show that the unlock times are diverse in each cluster by applying a non parametric pairwise comparison test. We used Kruskal-Wallis H-test followed by Dunn post-hoc comparisons and applied it to the unlock state data, using the clusters as sample groups. A 100% null hypothesis rejection rate (i.e. none of the samples varies significantly from any of the others) with $H = 67097.67$ and $p < 0.01$ suggests that the lock-states are not from the same population for each cluster.

B. Comparing against high-level features

To show the additional utility from considering user-app interaction behavior to characterize sessions, we examine how usage sessions could instead be grouped from the distributions of high-level features exclusively. Firstly, Table III shows that the distributions of individual features have long-tails with similar means and medians for most features. Splitting the distribution to group sessions using the range of the distribution results in most sessions being contained within a single group. For example, ~96% of sessions are placed within the same group for session length. Equally, splitting the distributions into tertiles, quartiles, or quintiles results in a high degree of similarity between most groups of sessions. This suggests that additional granularity is necessary to capture notable characteristics of usage and that the high-level features are not a suitable proxy for user-app interaction behaviour (as shown in the example in Figure 1).

To examine this further, Figure 4 shows how high-level features correlate to one another and what TF-IDF cluster each session is assigned to. Importantly, it shows that the TF-IDF clusters overlap and span across the distributions of high-level features, both individually and in pairs. This highlights that

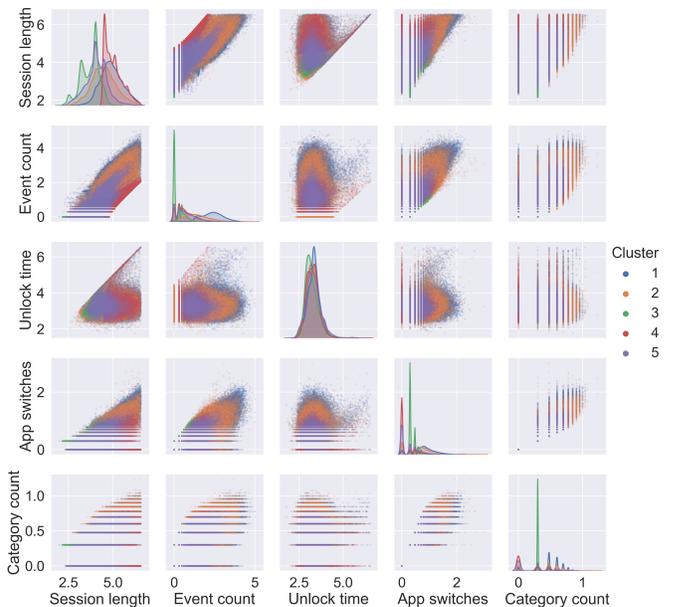


Fig. 4: Pairwise comparison of correlations between high-level features of sessions and TF-IDF clusters (Log-Log).

high-level features do not provide a suitable proxy for user-app interaction activity and that observing this granularity of behaviour is useful. We demonstrate this further by repeating the clustering process discussed in section III (with $k=5$) using a vector of all high-level features to represent a session, rather than TF-IDF scores of the lower-level features. Figure 3 shows how the clusters created by high-level features overlap poorly with the TF-IDF clusters by fixing the individual sessions in the same position as Figure 2.

V. CONCLUSIONS AND FUTURE WORK

The proposed methodology shows that clusters with similar high-level features can be substantially different at the event level, indicating the diversity of smartphone usage. It avoids potential issues of bias from using raws by utilising TF-IDF to increase the impact of significant events. We have analysed a rich dataset, which is unique in the level of user events that have been captured. Our results show that high-level features on their own are not sufficient to accurately group sessions that are indicative of the users cognitive goals. This is an important result, as previous analysis in the literature largely uses sessions defined by time between screen or application events [14, 7]. With regards to RQ1, our results confirm that clusters based only on high level features may misrepresent the commonality between smartphone sessions.

In terms of RQ2 this means that the additional information present when capturing low-level events is a useful tool to infer more about a session beyond how active a user was. We were able to capture and describe 5 different types of use that would not be transparent with previous techniques. In addition, these clusters represent usage that applies to almost all users,

with 84 out of 86 users in the dataset showing at least one session in each cluster.

A. Future work

Although the dataset used is extremely detailed in terms of events captured, a larger cohort of users may identify further session types beyond the five identified here. Further experiments with k (or alternative clustering techniques) may also identify additional coherent sub-types of sessions. However, as the data was collected in an in-the-wild environment, we are confident that these session types are representative of natural smartphone use.

Additionally, events are not singular actions that have no context to each other. Sequencing is a large enough interest point that it finds its way into other research areas of usage (such as application switching). TF-IDF does not take into account in which order words are structured in a document, therefore the current findings have room for more work here. To enable adding context to these events it would be possible to use n-gram as a combining factor. The continuity between events this provides might identify other clusters or groups of usage that are not possible to capture when considering all terms independently.

Currently all users are taken into account equally without specifically targeting high/low usage or any other factors that might contribute in how a user is interacting with their phone. Instead of going with this catch all approach it would be interesting to extend the low-level analysis with a more focused per-user approach. Finally, another aspect of future work would be to explore the predictability of the 'types' of session using the first few seconds of usage. Doing so would enable the ability to deliver interventions that may promote or disrupt the predicted usage trajectory - e.g. for giving the user notice of what other similar tasks are outstanding.

REFERENCES

- [1] C. Tan, Q. S. Liu, E. Chen, and H. Xiong, "Prediction for Mobile Application Usage Patterns," ser. Nokia MDC workshop '12, 2012.
- [2] H. Falaki, R. Mahajan, S. Kandula, D. Lyberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *Proc. MobiSys '10*. ACM, 2010, pp. 179–194.
- [3] T. Soikkeli, J. Karikoski, and H. Hammainen, "Diversity and End User Context in Smartphone Usage Sessions," in *Proc. NGMAST 2011*, 2011, pp. 7–12.
- [4] S. Zhao, J. Ramos, J. Tao, Z. Jiang, S. Li, Z. Wu, G. Pan, and A. K. Dey, "Discovering different kinds of smartphone users through their application usage behaviors," in *Proc. UbiComp '16*. ACM, 2016, pp. 498–509.
- [5] B. McKenzie and A. Cockburn, "An Empirical Analysis of Web Page Revisitation," in *Proc. HICSS '01*. IEEE Computer Society, 2001, p. 5019.
- [6] T. Yan, D. Chu, D. Ganesan, A. Kansal, and J. Liu, "Fast app launching for mobile devices using predictive user context," in *Proc. MobiSys '12*. ACM, 2012, pp. 113–126.
- [7] A. Oulasvirta, T. Rattenbury, L. Ma, and E. Raita, "Habits make smartphone use more pervasive," *Personal and Ubiquitous Computing*, vol. 16, no. 1, pp. 105–114, 2012.
- [8] S. L. Jones, D. Ferreira, S. Hosio, J. Goncalves, and V. Kostakos, "Revisitation analysis of smartphone app use," in *Proc. UbiComp '15*. ACM, 2015, pp. 1197–1208.
- [9] J.-M. Kang, S.-s. Seo, and J. W.-K. Hong, "Usage pattern analysis of smartphones," in *Proc. APNOMS'11*, Sep. 2011, pp. 1–8.
- [10] D. Ferreira, J. Goncalves, V. Kostakos, L. Barkhuus, and A. K. Dey, "Contextual experience sampling of mobile application micro-usage," in *Proc. MobileHCI '14*. ACM, 2014, pp. 91–100.
- [11] L. D. Turner, R. M. Whitaker, S. M. Allen, D. Linden, K. Tu, J. Li, and D. Towsley, "Evidence to support common application switching behaviour on smartphones," *Royal Society Open Science*, vol. 6, no. 3, p. 190018, 2019.
- [12] L. D. Turner, S. M. Allen, and R. M. Whitaker, "The influence of concurrent mobile notifications on individual responses," *International Journal of Human-Computer Studies*, vol. 132, pp. 70–80, 2019.
- [13] N. Banovic, C. Brant, J. Mankoff, and A. Dey, "ProactiveTasks: the short of mobile device use sessions," in *Proc. MobileHCI '14*. ACM, Sep. 2014, pp. 243–252.
- [14] C. Jesdabodi and W. Maalej, "Understanding usage states on mobile devices," in *Proc. UbiComp '15*. ACM, 2015, pp. 1221–1225.
- [15] R. Kumar, V. V. Phooha, and A. Serwadda, "Continuous authentication of smartphone users by fusing typing, swiping, and phone movement patterns," in *Proc. BTAS'16*, 2016, pp. 1–8.
- [16] C. Epp, M. Lippold, and R. L. Mandryk, "Identifying emotional states using keystroke dynamics," in *Proc. CHI '11*. ACM, 2011, pp. 715–724.
- [17] B. Noë, L. D. Turner, D. E. J. Linden, S. M. Allen, G. R. Maio, and R. M. Whitaker, "Timing rather than user traits mediates mood sampling on smartphones," *BMC Research Notes*, vol. 10, no. 1, p. 481, 2017.
- [18] Z. X. Jin, T. Plocher, and L. Kiff, "Touch Screen User Interfaces for Older Adults: Button Size and Spacing," ser. LNCS, C. Stephanidis, Ed., vol. 4554. Springer, 2007, pp. 933–941.
- [19] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, Jan. 1988.
- [20] J. Heer and E. H. Chi, "Separating the swarm: categorization methods for user sessions on the web," in *Proc. CHI '02*. ACM, 2002, pp. 243–250.
- [21] A. Gisbrecht and B. Hammer, "Data visualization by nonlinear dimensionality reduction," *WIREs Data Mining and Knowledge Discovery*, vol. 5, no. 2, pp. 51–73, 2015.