Are You There? Identifying Unavailability in Mobile Messaging

Pranut Jain
University of Pittsburgh
Pittsburgh, PA, USA
pranut@cs.pitt.edu

Rosta Farzan
University of Pittsburgh
Pittsburgh, PA, USA
rfarzan@pitt.edu

Adam J. Lee
University of Pittsburgh
Pittsburgh, PA, USA
adamlee@cs.pitt.edu

ABSTRACT

Delays in response to mobile messages can cause negative emotions in message senders and can affect an individual's social relationships. Recipients, too, feel a pressure to respond even during inopportune moments. A messaging assistant which could respond with relevant contextual information on behalf of individuals while they are unavailable might reduce the pressure to respond immediately and help put the sender at ease. By modelling attentiveness to messaging, we aim to (1) predict instances when a user is not able to attend to an incoming message within reasonable time and (2) identify what contextual factors can explain the user's attentiveness—or lack thereof—to messaging. In this work, we investigate two approaches to modelling attentiveness: a general approach in which data from a group of users is combined to form a single model for all users; and a personalized approach, in which an individual model is created for each user. Evaluating both models, we observed that on average, with just seven days of training data, the personalized model can outperform the generalized model in terms of both accuracy and F-measure for predicting inattentiveness. Further, we observed that in majority of cases, the messaging patterns identified by the attentiveness models varied widely across

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312893

KEYWORDS
Messaging; Attentiveness; Availability; Personalized Models.

users. For example, the top feature in the generalized model appeared in the top five features for only 41% of the individual personalized models.

INTRODUCTION

Past research has shown that users are highly attentive to messaging [5]. In cases where a recipient may not be able to attend to incoming messages immediately, they feel social pressure due to common messaging behavior expectations [11]. A survey [7] of messaging users awaiting response has shown that while more than 20% of the senders deem a recipient as 'is busy', about 15% also speculated that the recipient 'is pointedly ignoring me' (15.4%) or the recipient 'maybe in trouble' (5.7%). These speculations can cause a range of negative emotions in the sender and can affect social relationships [4]. Recipients often feel obligated to respond even only if to communicate unavailability and also make an attempt to justify delays in responding with circumstantial reasons based on their engagements and state of activities, e.g., 'I was on a call' [14].

We envision a messaging assistant that can identify when a user is unavailable and respond on their behalf with relevant information to explain their unavailability to the sender. This may alleviate some of this social pressure to respond and allow users to engage with messaging applications only during periods of true availability. By modelling attentiveness based on past messaging behavior, it may be possible to identify which contextual factors affect a user's availability to attend to messaging applications. An individual's *responsiveness* to notifications is typically difficult to predict and is based upon additional factors such as message content and context, and the relationship between the sender and receiver [9]. Thus, in this work, we focus on *attentiveness* to messaging, which is the degree to which a user is paying attention to incoming instant messages [11].

Attentiveness vs. Responsiveness to messaging: A user is *attentive* to messaging if they are aware of an incoming message and any details about it [11]. Modelling attentiveness to messaging deals with predicting whether or not the user is going to attend to an incoming message within a few minutes. A user can attend to an incoming message by accessing the notifications drawer, opening the application which generated the notification, or accessing the message on another device [5]. A user is *responsive* to a message if they respond to an incoming message within a certain amount of time [2]. Modelling responsiveness typically requires deep consideration of message content and relationship context [9].

One way to infer an individual's attentiveness is from cues provided by messaging applications. WhatsApp, Facebook, and Skype share the 'Last Seen' time, 'Read Receipts,' and Availability status (online, away) with communication initiators. However, it has been shown that these cues are not always good indicators of recipient's availability and can raise privacy concerns [4, 7, 11]. Previous work [11] proposed to share a user's predicted attentiveness level (high or low) as a cue before communication is initiated. The prediction was based on a generic model from the data of 24 participants collected over two weeks. While this approach may prevent a user from initiating communication during inopportune moments, it faces limitations in terms of our goals of facilitating a more effortless and efficient communication: (1) the user might not re-initiate communication later and the receiver may miss potentially important information; (2) incorrectly predicting a user as 'high attentive' when they are not might increase the expectation of a fast response; and (3) it has previously been shown that smartphone usage varies by demographics [1]. Users may exhibit different messaging behaviors in similar contexts; for instance, one user might be attentive to messaging while commuting whereas another user might not be. Aggregating data from multiple users to form a single model in these cases may not yield optimal results.

Thus, to address our goals of facilitating more efficient messaging communication, we aim to investigate modelling approaches that can accurately identify recipients' *unavailability* and use their contextual information to explain it. We investigate two approaches to model a user's attentiveness: (1) a *general approach* in which a generic model is built from the aggregate data of a group of individuals [5, 11] and (2) a *personalized approach* where individual models are built for each user. While in some tasks, personalized models have been shown to outperform a general model, this is not always the case [8]. Further, we investigate the data requirements associated with training accurate personalized models of attentiveness as different tasks have shown different data requirements [6, 15]. Our results show that on average, with seven days of training data, the personalized model outperforms the general model, confirming that a personalized model can capture a user's messaging behavior in varying contexts more accurately.

Dataset Used: The dataset we used for this work is from a study [10] that was aimed at identifying opportune moments when a user can engage with the contents of a mobile notification. The dataset captured a variety of events such as incoming notifications, phone calls and sensor events (e.g., screen status, noise level etc.) on a user's smartphone.

From this dataset, we extracted notification events by WhatsApp messenger since they made up 91% of all notifications posted by communication category applications. That data included 1,375,359 instances of WhatsApp notifications from 274 participants in the dataset spanning an average time period of 3 weeks.

Feature set: Our feature set included (1) current state of the device, e.g., screen status, or foreground running app; (2) last access-times, e.g., time since an application was opened, or since the device was last unlocked; (3) usage behavior in the last 60 minutes, e.g., number of WhatsApp notifications, or total battery drain; (4) usage behavior in the current day, e.g., percentage of time spent at home/work/commuting, or total data transmitted. In total, we had 72 features.

Target Variable: Our target variable to predict is whether or not the user attended to the notification within a certain threshold of time.
A notification can be attended by (1) accessing the notification drawer; (2) opening the application which created the notification; (3) accessing the notification on another device [5].
On average, the median time to attend a WhatsApp notification was 5.10 minutes which is the threshold value we used in our modeling [11].

MODELLING

General Model

Our general model follows the approach used by Pielot at al. [11] in modelling attentiveness to mobile messaging based on a general model that aggregates data from all individuals in the study.

To create the model, we utilized a gradient boosting decision tree approach XGBoost [3]. XGBoost has been shown to achieve better performance in multiple classification tasks including interruptibility prediction [10]. In our testing as well, it outperformed approaches like Random Forests and Logistic Regression. Setting the parameters 'max_depth' to 5 and 'min_child_weight' to 20 while leaving other parameters to default gave us the best performance during the tuning process. Using 10-fold grouped cross-validation technique with unique users within each fold of training and testing data, we achieved an average accuracy of 72.28% and F-measure value for inattentive class of 0.651. This type of evaluation allows us to get an estimate of how the model will perform for a new user for whom the model has not seen any past behavioral data. The top 10 features ranked according to the gain provided to the model are shown in the Figure 1. The 'timeSince' features represent the time that has passed since that event occurred. For instance, 'timeSinceLastOpenApp' represents the time passed (in milliseconds) since an application was last opened at the time of an incoming message. Similarly, 'Screen_Value' feature represents the screen state (on, off or unlocked) and 'Charging_Value' represents whether the phone was charging at the time of an incoming message.

We also investigated the effect of adding demographics information like age, gender and locale to the feature set to assess if the general model would benefit from this information. We observed an insignificant improvement in both accuracy ($p = 0.189$) and F-measure (inattentive) ($p = 0.099$).

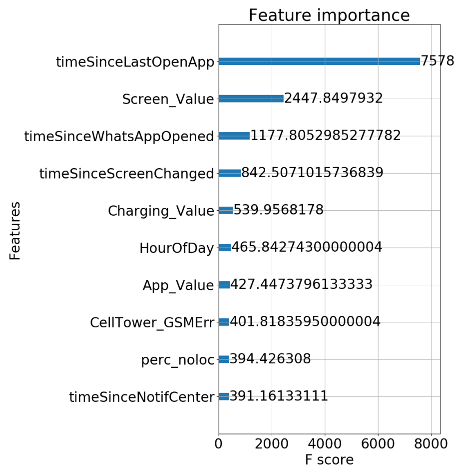

Figure 1: Top ranked features in the General Model

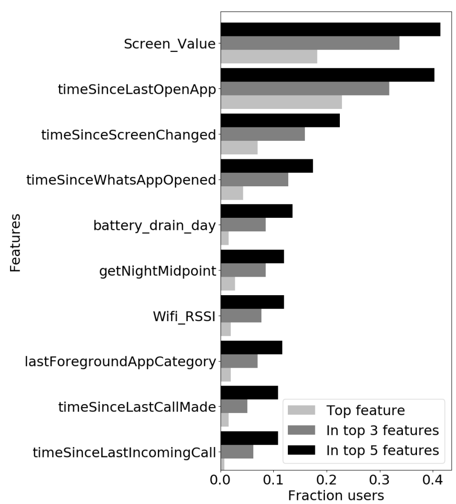

Figure 2: Top features for individual user models

Personalized Modelling Approach

It has been shown that an individuals' characteristics such as demographics are associated with different smartphone usage patterns [1]. Previously, personalized models for tasks like call-availability prediction [6] and interruptibility prediction [15] have been shown to outperform generic models. Capitalizing on this prior trend of research, we investigated the potential gain of a personalized model of prediction based on users' own prior data in comparison to prediction using all available data from the population.

For creating personalized models, we again used XGBoost [3] with default parameters and boosting iterations set to 20. To evaluate individual models, we used 10-fold grouped cross-validation technique. Messaging notifications are time-ordered, which means that in cases where a user is engaged in a series of back and forth interactions (i.e., instant messaging sessions [2]), some notifications may not be independent of each other. Further, new notifications may be generated for old unattended messages when a new message arrives. Thus, randomized cross-validation tends to over-estimate the model performance, whereas sequential-validation underestimates it [13]. Since there is no fixed time interval between notifications, in order to divide notifications into sessions of messaging, we added a session identifier to the notifications that arrived close to each other (within 15 seconds). Thus, when evaluating each model using cross-validation, we ensured that notifications were not split within a session and only across the sessions in each fold. We achieved mean accuracy score of 84.21% and mean F-measure for inattentive class of 0.744. This result suggests a significant improvement when considering a personalized model over a generic model.

Figure 2 shows the top ranked features for individual models. The plot shows the fraction of users for whom a feature was (1) Top ranked feature (2) In Top 3 features and (3) In Top 5 features. For instance, the feature 'timeSinceLastOpenApp' which denotes the time passed since any application was last opened is the top ranked feature for about 23% of individual models, in the top 3 features for 32% and in the top 5 features for 41% of the personalized user models.

Personalized Data Requirements

In the last section, we showed that the personalized modelling approach outperforms the general modelling approach. An important concern with personalized models can be the lack of initial training data for a new user, which can lead to suboptimal performance, even in comparison with a general model [8]. Such problems can be addressed by bootstrapping the model of a new user with a general model or a model formed from data gathered from similar users [15] until sufficient data becomes available to build a personalized model.

To investigate how much data will be sufficient for a personalized model to outperform the generic model, we assessed the individual models with gradual increase of the training data, in increments of

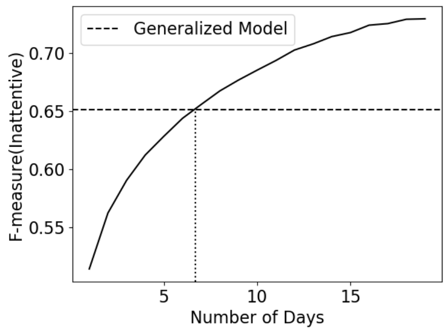

Figure 3: Number of days of training data and F-measure (inattentive)

days. For each user, we split the available data in the proportion of p/d, where d is the number of days represented in that user's data and p is the number of days to be used for training, which was varied from 1 to $(d-1)$. The rest of the data was used as testing data. We followed the session-based evaluation approach mentioned in the last section. For each user, the process was repeated 10 times and the results were averaged.

Figure 3 presents the change in F-measure for the inattentive class as the more days of training data is added. As the number of days of training data is increased, the average performance goes up. After using 7 days of training data, the personalized modelling approach outperforms the generalized approach and with 16 days of training data the model performance stabilizes.

DISCUSSION AND FUTURE WORK

We showed that with sufficient training data, personalized modelling outperforms the general approach in both accuracy and F-measure when predicting unavailability. Further, the top features identified in the general model appeared in the top five features for only 41% of the individual personalized models. This result confirms that messaging behavior and attention to mobile messaging varies across individuals, and that a general model cannot accurately explain the variety of behaviors. Thus, utilizing a personalized model of attentiveness not only achieves higher accuracy, but can also provide insight regarding the contextual features leading to a user's unavailability.

With a personalized modelling approach, it will also be possible to update the model to adapt to changes in a user's environment and messaging behavior. We have shown that after utilizing 16 days of training data, personalized model performance starts to converge. Thus, a personalized model can be kept updated by retraining periodically with the user's last 16 days of usage data. Furthermore, with advances in mobile technologies and improving processing power of mobile devices, it will be possible to locally train models on users' devices [12] without needing to send their data to a remote server for processing, thereby reducing potential privacy concerns.

The current work is the first step in building a messaging assistant to support individuals in times of unavailability and reduce the stress associated with perceived obligation. We were able to model user's messaging behavior and predict their attentiveness state with high accuracy. By interpreting these attentiveness models we can determine which contextual factors at the time of an incoming message had the most weight in pushing the prediction of the model towards the inattentive state. Using such information, an automated response can vary from a simple message of 'Busy at the moment' to more elaborated explanation such as explaining individuals' engagement based on their calendar or current activity (e.g., 'Busy in a meeting' or 'Busy driving'). In constructing such automated responses, it is essential to consider users' privacy and nuances of social relationships. The ability to develop a computational model to accurately assist individuals in mobile messaging while respecting and accommodating their privacy can notably support mobile-based communication.

Acknowledgements The authors wish to thank Martin Pielot for providing the data set used in this work. This work was supported in part by the National Science Foundation under awards CNS–1253204 and CNS–1814866.

REFERENCES

[1] Ionut Andone, Konrad Błaszkiewicz, Mark Eibes, Boris Trendafilov, Christian Montag, and Alexander Markowetz. 2016. How age and gender affect smartphone usage. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 9–12.

[2] Daniel Avrahami and Scott E Hudson. 2006. Responsiveness in instant messaging: predictive models supporting interpersonal communication. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 731–740.

[3] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 785–794.

[4] Karen Church and Rodrigo De Oliveira. 2013. What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*. ACM, 352–361.

[5] Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 1–5.

[6] Robert Fisher and Reid Simmons. 2011. Smartphone interruptibility using density-weighted uncertainty sampling with reinforcement learning. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, Vol. 1. IEEE, 436–441.

[7] Roberto Hoyle, Srijita Das, Apu Kapadia, Adam J Lee, and Kami Vaniea. 2017. Was my message read?: Privacy and Signaling on Facebook Messenger. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3838–3842.

[8] Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. 2003. Predicting human interruptibility with sensors: a Wizard of Oz feasibility study. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 257–264.

[9] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My phone and me: understanding people's receptivity to mobile notifications. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 1021–1032.

[10] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond interruptibility: Predicting opportune moments to engage mobile phone users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 91.

[11] Martin Pielot, Rodrigo de Oliveira, Haewoon Kwak, and Nuria Oliver. 2014. Didn't you see my message?: predicting attentiveness to mobile instant messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3319–3328.

[12] Sujith Ravi. 2017. On-Device Machine Intelligence. https://ai.googleblog.com/2017/02/on-device-machine-intelligence.html.

[13] David R Roberts, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40, 8 (2017), 913–929.

[14] Amy Voida, Wendy C Newstetter, and Elizabeth D Mynatt. 2002. When conventions collide: the tensions of instant messaging attributed. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 187–194.

[15] Fengpeng Yuan, Xianyi Gao, and Janne Lindqvist. 2017. How busy are you?: Predicting the interruptibility intensity of mobile users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5346–5360.