Predicting Smartphone Users' General Responsiveness to IM Contacts Based on IM Behavior

Hao-Ping Lee
National Chiao Tung University
Hsinchu, Taiwan
dimension4.cs03@nctu.edu.tw

Yu-Lin Chung
National Tsing Hua University
Hsinchu, Taiwan
charles5j21z@gmail.com

Tilman Dingler
The University of Melbourne
Melbourne, Australia
tilman.dingler@unimelb.edu.au

Chia-Yu Chen
National Chengchi University
Taipei, Taiwan
chiayuchen.tw@gmail.com

Chih-Heng Lin
Kuan-Yin Chen
National Chiao Tung University
Hsinchu, Taiwan
cowbon.cs03@g2.nctu.edu.tw
kevchentw.cs03@g2.nctu.edu.tw

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

MobileHCI '19, October 1–4, 2019, Taipei, Taiwan.

ACM ISBN 978-1-4503-6825-4/19/10 $15.00.
https://doi.org/10.1145/3338286.3344387

Abstract
History of conversations through instant messaging (IM) contains abundant information about the communication patterns of the dyad, including conversation partners' mutual responsiveness to messages. We have, however, not seen many examinations of using such information in modeling mobile users' responsiveness in IM communication. In this paper, we present an in-the-wild study, in which we leverage participants' IM messaging logs to build models predicting their general responsiveness. Our models based on data from 33 IM user achieved an accuracy of up to 71% (AUROC). In particular, we show that 90-day IM-communication patterns, in general, outperformed their 14-day equivalent in our prediction models, indicating better coherence between long-term IM patterns with their general communication experience.

Author Keywords
Mobile notifications; mobile receptivity; ESM; machine learning.

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; H.1.2 [Models and Principles]: User/Machine Systems

Introduction

Communicating via instant messaging (IM) has become a common daily practice among smartphone users. While this contributes to feelings of connectedness, it can also be a source of distraction for the recipient and create social pressure [6]. Researchers have explored using context information extracted from smartphone users' phones to predict their attentiveness to mobile IM messages [2]; moments when they will feel interruptible; and opportune moments to which mobile notifications can be deferred [3, 8]. Lee *et al.* [5] further consider how the relationship with the sender affects the user's receptivity to IM notifications. While these research studies have provided us with insights of recognizing responsiveness because of their relatedness to responsiveness, IM message-history logs (hereafter abbreviated as "IM logs") may give us a more fine-grained picture of individual sender-receiver communications. Unfortunately, such resources have seldom been utilized to model mobile users' IM responsiveness.

This paper examines the effectiveness of using IM log data to predict mobile IM users' **General Responsiveness** toward a specific contact, i.e. how responsive users are to a specific contact's IM messages in general. We consider two types of General Responsiveness in our research: 1) users' general perception of their own responsiveness to their specific contacts, 2) an average of users' in-situ responsiveness to the contacts' IM messages. We refer to the former *Perceive General Responsiveness*. For the latter, we consider two ways of obtaining in-situ responsiveness data to generate such an average: in-situ self-reported responsiveness via Experience Sampling Method(ESM), referred to as *average ESM self-report responsiveness*, and responsiveness extracted from IM Log, referred to as *average IM log actual responsiveness*. Our research question is: *To what extent can IM log data be used to predict these three types of General Responsiveness of mobile users?* We conducted a 4-week in-the-wild study, in which we collected these three types of General Responsiveness measures from participants, and extracted features from their IM logs to predict these measures respectively. We show that IM logs are effective in predicting mobile users' General Responsiveness to IM contacts. Furthermore, 90-day IM-communication patterns, in general, outperformed their 14-day equivalents; a combination of these two time-windows create almost equivalent predictions with using only 90-day IM-communication patterns.

Data Collection

As mentioned earlier, we consider three measures of General Responsiveness: *perceive general responsiveness*, *average ESM self-report responsiveness* and *average IM log actual responsiveness*. Each measure was treated as a binary outcome, i.e., responsive/ non-responsive, in the prediction models. Below, we explain how we collected responsiveness data from participants.

Collecting Perceive General Responsiveness

Study participants were asked to self-report their perceived general responsiveness to twenty IM contacts via contact questionnaires. The responsiveness question was: *In general, how responsive are you to his/ her notifications after seeing it?"*. A user was defined as being responsive to a contact's IM messages if s/he answered *"I usually respond immediately" or "I usually respond in several minutes."*

Collecting In-situ Responsiveness

We considered two types of in-situ responsiveness from which we derive an average measure. The first was ESM self-report responsiveness, where participants reported how fast they responded to a sampled IM notification via ESM. A user was defined as responsive to a message

Table 1: List of features related to IM pattern

IM Communication Pattern	
Intensity	{AVG,MIN,MAX,STD} {#,length} total msg per day
	MAX # msg in an hour
	MAX # msg in an hour / AVG # msg per hour
	{#,length} incoming msg / total msg
Regularity	# days have msg
	{AVG,MIN,MAX,STD} # hours have msg per day
Contact Tendency[a]	total {#,length}
	{#,length}incoming msg / total msg
	# days have msg
	AVG # hours have msg per day
	MAX # msg in an hour
Temporal Tendency	{#,length}{times of days, weekdays,weekends} msg / total msg
Session Character-istics[b]	time from the last session (minute)
	duration (minute)
	msg count
	turn count
	character count
	msg-per-minute
	msg-per-turn
	characters-per-msg
	MIN gap duration (sec)
	MAX gap duration (sec)
	AVG gap duration (sec)

[a]These features were calculated in contact personalized z-scores.
[b]These features were calculated in {AVG,MIN,MAX,STD}.

if s/he agreed with either of the following two ESM statements: *"I will respond/responded to it immediately"* or *"I will respond/responded to it in several minutes"*. We averaged these self-reported responsiveness to generate an *average ESM self-report responsiveness*. Specifically, we yielded the ratio of his/her being responsiveness to the contact according to the his/her responses. The ratio higher than 0.5 indicated that the participant was more often to be responsiveness to the contact's incoming messages, and thus was defined as *responsive* in **average ESM self-report responsiveness** to that contact. Conversely, a lower result led to the participant being defined as *non-responsive*.

We extracted IM log actual responsiveness from IM logs. A participant was defined to be *responsive* to a sampled message if the elapsed time between when the message was received and when the user responded to it was within a certain time threshold. An appropriate time threshold should: 1) distinguish between high and low IM log actual responsiveness. 2) not result in a large discrepancy between ESM self-report responsiveness and IM log actual responsiveness because we need to use both for predicting responsiveness. Time threshold is tested from t = 0 (second) and is iterated with a 1-second interval. We find a saturation point for matching actual IM log and self-report ESM responsiveness at around 10 minutes (t = 645), where the consistency of the two is 78%. In the end, we select 645 seconds as a time threshold to distinguish actual responsiveness/ non-responsiveness. We then compute the ratio of the responses being actual responsiveness of each contact, and define that a participant is responsive to contact in terms of **average IM log actual responsiveness** when it is higher than 0.5.

User Study
Details of the study in this work has been reported in [5]. Below we briefly summarize it for the completeness of the paper. In total, 34 participants completed the experiment. All of them participated for at least four weeks (M=37.5), in which each of them completed 20 contact questionnaires, and contributed in total 4,570 ESM responses. The demographic included 20 students and 14 non-students, 17 males and 17 females, aged 20 to 50 (M=25.33). All of them were active users of Facebook Messenger and/or Line Messenger, the two most popular IM apps in Taiwan.

IM Conversation Features Selection

Among the 33 participants who provided IM logs for their contacts in the study, we collected message histories of 569 participant-IM contact pairs, with an average 17.24 contacts per participant. For each of the participant-contact pair, we extracted 84 features that describe the IM communication patterns of the dyad, which are used to predict participant's General Responsiveness to that contact.

The features include five facets of IM communication patterns, as detailed in Table 1, including: *intensity* (e.g., total number of messages), *regularity* (e.g., days with messages), *temporal tendency* (e.g., the proportion of messages that arrive in the evening), *contact tendency*, which measured the differences in communication patterns with a given contact as opposed to the participant's 5 frequent contacts with most messages, and *session characteristics*, which followed the concept advanced by Issacs *et al.* [4] that messages arrived within 5 minutes of a message were in the same communication session, and adopted features suggested by Avrahami *et al.* [1], including turn (e.g. one turn comprises consecutive messages from only user or sender) and gap (time between two turns) to derive features reflecting their general conversational sessions.

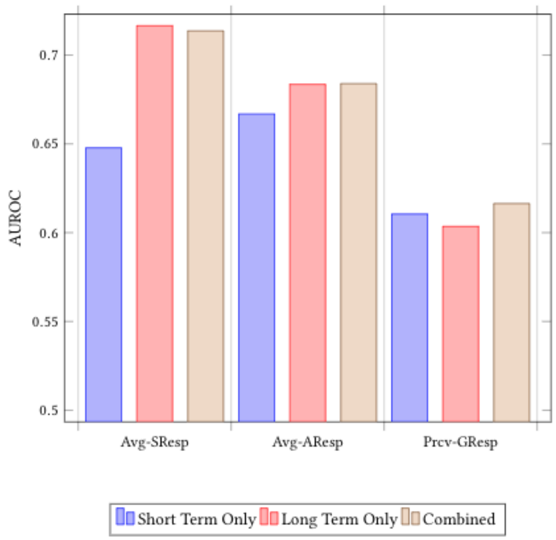

Figure 1: General Responsiveness prediction with different feature configurations.

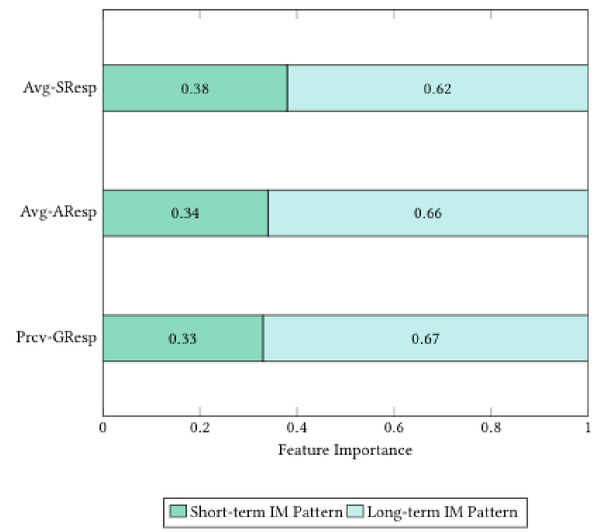

Figure 2: Feature exploration, General Responsiveness prediction using combined feature configuration.

Inspired by Reinhardt *et al.* [7], for all of the features we generated two IM Communication Pattern feature sets for each participant-contact pair from a message history of 14 days and 90 days, to represent short-term and relatively long-term communication patterns, respectively. The 14 and 90 days were counted backward from the day the participant finished the contact questionnaire for that contact.

General Responsiveness Prediction

We use IM log data to build prediction models of the participants' General Responsiveness. We selected three widely used classifiers: 1) Random Forest (RF) 2) SVM with Radio Basis Functions kernel, and 3) Extreme Gradient Boosting (XGBoost). Each classifier was subject to a gridsearch to select hyper-parameters for generating the best performance. For each hyper-parameter set we ran a 10-fold cross validation to evaluate the performance. We utilized Area Under the Receiver Operating Characteristics (AUROC)[1] as our models' evaluation metrics, due to its more accurate evaluation of imbalanced datasets. We achieved the best performance with Random Forest in predicting average IM log actual responsiveness, and with XGBoost for all the other models, respectively.

In building the predictive models, we explore three configurations of communication patterns: 1) long-term pattern (only IM patterns of 90 days), 2) short-term pattern (only IM patterns of 14 days), and 3) a combination of the two. Three General Responsiveness measures were predicted:

[1] https://en.wikipedia.org/wiki/Receiver_operating_characteristic

Table 2: Performance Results and Top Features in General Receptivity Prediction

Responsiveness Measure	AU-ROC	Prec-ision	Re-call	F1	Kappa	Top Features	Feature Category	Feature Importance
Average ESM Self-report	0.71	0.79	0.96	0.87	0.17	avg_session_turn_count_90	LT	0.039
						avg_session_minimum_gap_90	LT	0.0374
						max_daily_msg_length_14	ST	0.0354
						avg_session_message_count_90	LT	0.0298
						std_session_message_per_turn_90	LT	0.0287
Average IM Log Actual	0.68	0.7	0.87	0.77	0.19	min_session_interval_gap_90	LT	0.0143
						std_session_maximum_gap_90	LT	0.0139
						std_session_message_per_turn_90	LT	0.0135
						contact_zscore_ratio_of_incoming_msg_90	LT	0.0134
						avg_session_average_gap_90	LT	0.0125
General Percieved	0.62	0.64	0.72	0.68	0.16	ratio_afternoon_msg_90	LT	0.0139
						ratio_of_early_night_msg_length_90	LT	0.0138
						avg_daily_#_of_msg_90	LT	0.0134
						ratio_of_afternoon_msg_length_90	LT	0.0133
						contact_zscore_ratio_of_incoming_msg_length_90	LT	0.013

LT=Long-Term IM Pattern; ST=Short-term IM Pattern

average ESM self-report responsiveness (shown in the figures and hereafter as Avg-SResp), average IM log actual responsiveness (Avg-AResp), and perceived general responsiveness (Prcv-GResp).

The General Responsiveness prediction results are shown in Figure 1. Comparison of the three pattern configurations revealed that long-term patterns achieved better predictive performance than short-term ones for the two average responsiveness measures, i.e. Avg-SResp and Avg-AResp. Adding features of short-term patterns (i.e. the combination), the model achieved almost equivalent performance. In contrast, short-term patterns outperform long-term patterns only in predicting Prcv-GResp. Again, adding features of long-term patterns into the model did not increase the performance considerably. Notably, among our combined model's top 15 features having the highest importance, as shown in Table 2, 14 features were of long-term IM patterns. Features of those patterns also played a disproportionate role in contributing to the prediction, i.e., with a category maximum of up to 67%, as against no more than 38% for short-term patterns (see Figure 2). These results suggest that long-term IM patterns were more indicative of participants' General Responsiveness than short-term ones. Yet, short-term ones still provide useful information for prediction, i.e., about one third of the overall model contribution in the combination model.

Interestingly, the variable that was based on averaged ESM responses, Avg-SResp, were easier to predict than its counterpart version of General Responsiveness, Prcv-GResp. Moreover, whereas the top-ranked features for Avg-SResp included features related to session characteristics, Prcv-GResp did not include any of these features. This could imply that how a participant typically interacted with an IM contact in daily conversation was less indicative of participant's general perceptions of his/her own responsiveness than it was of measures averaged from in-situ experiences.

Summary

We show the feasibility of using IM history logs to predict mobile users' General Responsiveness to their contacts without intruding their content privacy and achieved the performance at best 71% AUROC. We believe the finding could inspire future research and IM services to consider IM logs as the sources of understanding mobile users' behavior on IM communication.

REFERENCES

1. Daniel Avrahami and Scott E. Hudson. 2006. Communication Characteristics of Instant Messaging: Effects and Predictions of Interpersonal Relationships. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*. ACM, New York, NY, USA, 505–514. DOI: `http://dx.doi.org/10.1145/1180875.1180954`

2. 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 (MobileHCI '15)*. ACM, New York, NY, USA, 1–5. DOI: `http://dx.doi.org/10.1145/2785830.2785840`

3. 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*. ACM, 181–190.

4. Ellen Isaacs, Alan Walendowski, Steve Whittaker, Diane J. Schiano, and Candace Kamm. 2002. The Character, Functions, and Styles of Instant Messaging in the Workplace. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work (CSCW '02)*. ACM, New York, NY, USA, 11–20. DOI: `http://dx.doi.org/10.1145/587078.587081`

5. Hao-Ping Lee, Kuan-Yin Chen, Chih-Heng Lin, Chia-Yu Chen, Yu-Lin Chung, Yung-Ju Chang, and Chien-Ru Sun. 2019. Does Who Matter?: Studying the Impact of Relationship Characteristics on Receptivity to Mobile IM Messages. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 526, 12 pages. DOI: `http://dx.doi.org/10.1145/3290605.3300756`

6. Martin Pielot, Karen Church, and Rodrigo De Oliveira. 2014. An in-situ study of mobile phone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 233–242.

7. Delphine Reinhardt, Franziska Engelmann, Andrey Moerov, and Matthias Hollick. 2015. Show Me Your Phone, I Will Tell You Who Your Friends Are: Analyzing Smartphone Data to Identify Social Relationships. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15)*. ACM, New York, NY, USA, 75–83. DOI: `http://dx.doi.org/10.1145/2836041.2836048`

8. 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 (CHI '17)*. ACM, New York, NY, USA, 5346–5360. DOI: `http://dx.doi.org/10.1145/3025453.3025946`