Distinguishing IM Communication Patterns with Relationship and Conversation Topics

Kung-Pai Lin*
bwboao.cs05@g2.nctu.edu.tw
Department of Computer Science, National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Hao-Ping (Hank) Lee*
hankhplee@gmail.com
National Taiwan University
Taipei, Taiwan
Department of Computer Science, National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Yu-Ling Chou
egghead70187@gapp.nthu.edu.tw
National Tsing Hua University
Hsinchu, Taiwan

Faye Shih
fshih@brynmawr.edu
Bryn Mawr College
Bryn Mawr, USA

Yung-Ju Chang
armuro@cs.nctu.edu.tw
Department of Computer Science, National Yang Ming Chiao Tung University
Hsinchu, Taiwan

ABSTRACT

Despite being characterized as constantly on and connected, IM users' responsiveness varies across different contacts. While research has shown that the relationship between conversation partners plays an important role in influencing their communication patterns, characterization of such patterns simply by using relationship information is limited [7, 8]. In this paper, we identify five distinct clusters of IM patterns using unsupervised learning derived from 46 users' conversation history. We show that the relationship category sufficed to characterize three clusters of communication patterns, but failed for the most active one. However, considering both relationship and topics would distinguish most communication patterns, including the most active one. This result suggests that future research on IM communication patterns should pay more attention to the topics in users' conversations.

ACM Reference Format:
Kung-Pai Lin, Hao-Ping (Hank) Lee, Yu-Ling Chou, Faye Shih, and Yung-Ju Chang. 2021. Distinguishing IM Communication Patterns with Relationship and Conversation Topics. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '21 Companion), October 23–27, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3462204.3481752

*Both authors contributed equally to this research.

CSCW '21 Companion, October 23–27, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8479-7/21/10…$15.00
https://doi.org/10.1145/3462204.3481752

1 INTRODUCTION

Instant messaging (IM) has become an essential channel for people to communicate about a variety of topics with their contacts in their day-to-day lives. However, while IM users are expected to be always-on and always-connected, and are presumably able to respond to messages soon and at any time [14], their responsiveness to messages nevertheless varies across not only contexts but also individual contacts [8]. Prior research has suggested that relationship plays a vital role in explaining varied responsiveness [2, 8]. Some researchers have also attempted to characterize IM users' communication patterns using relationship information. However, research also shows that conversation topic that IM users communicate about also plays an important role; it has been suggested that IM users may be more responsive to messages that are work-related [1]. Nevertheless, the main body of the literature mainly focuses on the impact of relationship-related factors on IM users' communication patterns between conversation partners; little attention has been paid to how typical conversation topics between partners affect their communication patterns. We argue that it may be necessary to consider both in order to better characterize IM users' communication patterns. For example, work-related topics may appear in conversations with various kinds of contacts, not only co-workers, but also family members and friends; but IM users may display different responsiveness to such messages from different kinds of contacts. By the same token, IM users may also chat about different topics with similar kinds of contacts, but are more responsive to those who typically talk about urgent business than those who only chitchat [4].

To fill the research gap, in our study, we examined whether considering both relationship and conversation topic better characterizes IM users' communication patterns than considering either of them on their own. We first utilized unsupervised learning on communication features extracted from IM conversation history logs from 46 users, totaling 689,418 messages, to identify five distinct communication patterns [7]. We then built logistic regression models using four different sets of features, 1) Topic, 2) Relationship, 3) Relationship & Topic, and 4) Relationship-Topic Combined, to characterize IM patterns, i.e. to predict whether a conversation pair belongs to one of the clusters.

Our findings show that the model with conversation topics alone achieved extremely poor performance in distinguishing all clusters of IM patterns. Second, we found that while relationship alone distinguished certain communication patterns well, it failed at the most active communication pattern. However, considering both topic and relationship effectively characterized all clusters except one. This model also achieved the highest average R^2 (0.67) in distinguishing all clusters. Finally, all sets of features failed to characterize the least active communication patterns, suggesting that other information is needed to distinguish this pattern from the others.

2 METHOD

We recruited 46 active IM users, 18 males, 27 females, and one who preferred not to answer, through a Facebook group intended for research-subject recruitment; 26 were students and 20 were employed across various fields. The age of the participants ranged from 20 to 56 years old (M = 26.1, SD = 6.7). We asked participants to select 20 of their IM contacts, for each of whom participants provided their relationship, interaction behaviors, IM practices, and conversation topics with the contact via a questionnaire. Participants also provided their chat logs on LINE and Facebook Messenger with the selected contacts. We compensated each participant with 750 NTD (roughly 27 USD) for participating in the study.

2.1 Measures

2.1.1 Relationship type. Participants chose their relationship types with selected contacts from the available options adopted from [7], including: significant others (SO), immediate family, parents, extended family, superior at work, subordinate at work, junior colleague, colleague, senior colleague, client, service provider, friend, acquaintance, classmate, and others. For our analysis, we grouped these relationship types into social, acquaintance, classmate, family, parent, relative, SO, superior at work, and work, reducing the number of similar relationship types.

2.1.2 Conversation Topic. Participants provided their most frequent conversation topics with the chosen contacts. According to [4, 5, 13], we organized conversation topics into three options — 1) job-related, 2) chitchat, and 3) coordination for specific purposes (e.g., hosting an event, finishing a group project). Participants were asked to choose all that applied, since conversations with a contact could contain multiple topics. For our analysis, we further categorized participants' responses into five categories — *Chat*, *Work*, *Purpose*(coordination for a specific purpose), *Mixed*, and *All*. Specifically, responses in which participants selected only either work, chat, or purpose, were assigned to that category; responses in which participants selected any two of the three options were assigned to the category *Mixed*; and responses in which participants selected all three options were assigned to the category *All*.

2.1.3 Relationship-topic combined type. We additionally created a category schema that represented a measure taking both relationship and conversation topic into account. This allowed us to examine if considering both would achieve better performance in characterizing communication patterns than considering either of them as a standalone measure. To create the categories, we first extracted relationship types that research has shown to have an impact on IM users' communication behaviors [4, 11]. They include 'SO,' 'family,' 'parents,' and 'superior at work.' We categorized responses into these categories if their corresponding options were chosen. These categories did not consider topics. The rest of the responses were categorized based on both the chosen relationship type and the topic, which were: 'pure chat' (chose social and chitchat) and 'pure work' (chose co-worker and work-related topic). These two categories thus consider the link between relationship type and topic. The rest of the responses that were not assigned a category were categorized by conversation topics: 'chat,' 'work,' 'purpose,' 'mixed' and 'all.' In other words, the category schema includes both relationship types and topics, and two additional categories that consider the relation between them.

2.2 Extracting IM Communication Features

We collected 963 conversation history logs, each of which was a conversation history between a participant and his/her selected contact. After removing logs that did not include any mutual conversation, 703 logs remained in our data analysis. Inspired by Reinhardt et al. [12], we extracted two communication features: intensity and regularity. We additionally included response time, following Lee et al. [7]. As a result, three features — total number of chats, average response time, and the number of days on which they chatted within a 90-day window [9, 12], which represented intensity, responsiveness, and regularity, were extracted from each conversation log. For each of these features, we also computed its "rank," which compared the current value of the feature with other selected contacts for the same participant (e.g. 1st out of 20: 100%; 20th out of 20: 5%). We added this information because an absolute value could be greatly influenced by individual differences, as some participants may tend to be responsive to all contacts whereas others tend to be the opposite. Rank information was not influenced by such individual differences. These six features were then used in building unsupervised models to distinguish IM communication patterns, as described below.

3 RESULT

3.1 Characterizing IM Communication Patterns

We used k-means [10], an unsupervised machine learning method, following prior work [7] to identify distinct IM communication patterns. This approach allows us to divide general IM behavior, while preventing us from only focusing on one side of the IM patterns e.g., total number of chats. We used the aforementioned six features to build the unsupervised learning model, and used the Elbow Method and Silhouette Method [6] to evaluate our results, shown in Fig. 1. We achieved the best results with k equal to five, resulting in five distinct clusters of IM communication patterns (see Table 1). Below

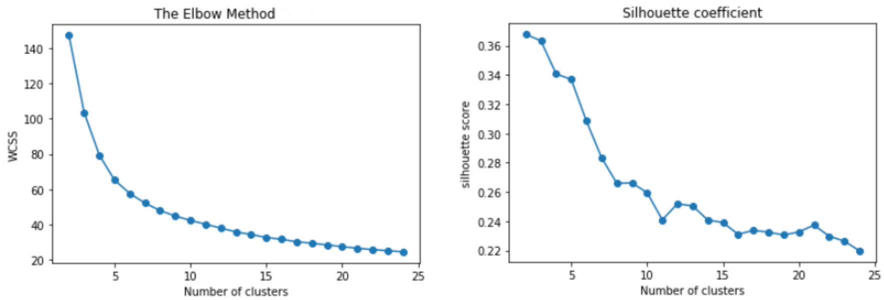

Figure 1: The two charts respectively show the result of WCSS(Within-Cluster Sum of Square) of different k values, and the result of silhouette coefficient of different k values. Combined together we chose k=5 with the value near the "elbow," which also has a higher silhouette coefficient.

Table 1: IM communication patterns of each cluster

Clusters	Barely-IM (N=161)	Promptly-IM-when-Necessary (N=137)	Regularly-and-Slowly-IM (N=146)	Sometimes-Actively-IM (N=145)	Actively-IM-All-the-Time (N=114)	Total (N=703)
Total msg # in 90 days	63.9	88.7	331.1	447.3	4,857.8	980.7
Average response time (mins)	154.4	15.1	84.6	25.3	37.4	67.2
days has msg in 90 days	7.4	3.9	24.2	13.5	61.4	20.2
Total msg #in 90 days ranked	24%	28%	64%	68%	92%	53%
Average response time ranked	23%	85%	29%	73%	61%	53%
days has msg in 90days ranked	29%	18%	70%	49%	90%	49%

Note. The numbers are the average of all pairs in each cluster

we briefly introduce the communication characteristics of the five clusters.

3.2 Five IM Communication Patterns

3.2.1 Barely-IM and Promptly-IM-when-Necessary. As shown in Table 1, the Barely-IM Cluster and the Promptly-IM-when-Necessary Cluster had the two least number of messages exchanged (63.9 vs. 88.7 messages), showing that participants rarely sent IM messages. However, the former had the slowest response pace (154.4 minutes) and the latter, interestingly, had the fastest response pace (15.1 minutes) among all the clusters. It is likely that the latter pattern took place more often between conversation partners who rarely exchanged messages, but when they did, they exchanged time-sensitive or important messages that needed to be responded to faster.

3.2.2 Regularly-and-Slowly-IM and Sometimes-Actively-IM. The Regularly-and-Slowly-IM Cluster and the Sometimes-Actively-IM Cluster were both average active. The former displayed lower responsiveness (84.6 minutes vs. 25.3 minutes) and exchanged fewer messages (331.1 messages vs. 447.3 messages), but exchanged messages on more days (24.2 days vs. 13.5 days). Therefore, it indicates more regular but slower message exchanges. The latter cluster indicates a more active but also more scattered conversations.

3.2.3 Actively-IM-All-the-Time. The Actively-IM-All-the-Time Cluster indicates the most active communication patterns, which has a significantly larger number of messages (4857.8), and the most frequent message exchanges (61.4 days). We considered this pattern to occur between partners with strong connections.

Table 2: Features used in each Regression Model

Regression Model	features include
Topic	conversation topic
Relationship	relationship type
Relationship & Topic	relationship type, conversation topic
Relationship-topic combined	relationship-topic combined type

3.3 Distinguishing Five Clusters using Relationship and Topic

In order to indicate how well the relationship and topic features can distinguish different clusters, we used logistic regression to predict cluster membership, i.e. predicting whether a conversation pair belonged to each of the five clusters or not, respectively, using the proposed features. We created four logistic regression models: the Topic Model, the Relationship Model, the Relationship&Topic Model and the Relationship-Topic Combined Model that used the aforementioned features (see Table 2). Specifically, we included IM users as a random variable to account for individual differences, and used the Pseudo R^2 value to indicate how well the models explain the variability in the odds that a pair of conversation partners

Table 3: Pseudo R^2 of each regression test

Regression model	Barely-IM (N=161)	Promptly-IM-when-Necessary (N=137)	Regularly-and-Slowly-IM (N=146)	Sometimes-Actively-IM (N=145)	Actively-IM-All-the-Time (N=114)	Overall
Topic	0.034	0.022	0.008	0.018	0.105	0.037
Relationship	0.044	0.901	0.836	0.845	0.153	0.556
Relationship & Topic	0.059	0.848	0.835	0.787	0.185	0.543
Relationship-topic combined	0.044	0.882	0.775	0.831	0.812	0.669

Note. The numbers are the average of all pairs in each cluster

belonged to that cluster. Thus, the higher the R^2 value, the more effectivly the model distinguishes that cluster from all the other clusters. The results are shown in Table 3.

Our regression results show that, first, the Topic Model had extremely low R^2 across all clusters, indicating that conversation topic alone poorly distinguished IM communication patterns. Second, the Relationship Model achieved high R^2 for three clusters, but it also poorly distinguished the Barely-IM and Actively-IM-All-the-Time clusters from the others. Furthermore, simply adding the effect of topic into the regression model did not improve the performance. The Relationship-Topic Combined Model, which used the proposed category that combined relationship and topic, was able to distinguish all clusters except Barely-IM. These results together suggested that relationship sufficed to characterize normally active communication patterns but not the least and the most active communication patterns. Taking both topic and relationship information into account was able to characterize the most active communication patterns. This model also achieved the highest average R^2 (0.67) in distinguishing the clusters, noticeably higher than the second highest (0.56). It is likely that, even with the same relationship type, participants would only actively exchange messages with those to whom they talked about certain topics.

4 CONCLUSION AND FUTURE WORK

We used the unsupervised learning method to identify five different IM communication patterns, and we have shown that while relationship type alone sufficed to characterize certain communication patterns, it failed to characterize the most active communication pattern. By using a category that considered both relationship and conversation topic, most communication patterns, including this highly active one, was successfully characterized. Overall, this work provided an exemplar that it is necessary and promising to consider relationship and conversation topics when characterizing IM communication patterns. However, to further improve the explanation of these communication patterns, future research should consider more factors that prior research has found to have an influence on IM responsiveness, such as the demographics of the chat partners, the closeness and communication expectations between the dyads [8], or the alert mode [3] that IM users are using on their phones.

REFERENCES

[1] Daniel Avrahami, Susan R. Fussell, and Scott E. Hudson. 2008. IM Waiting: Timing and Responsiveness in Semi-Synchronous Communication. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA, USA) *(CSCW '08)*. Association for Computing Machinery, New York, NY, USA, 285–294. https://doi.org/10.1145/1460563.1460610

[2] 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* (Banff, Alberta, Canada) *(CSCW '06)*. Association for Computing Machinery, New York, NY, USA, 505–514. https://doi.org/10.1145/1180875.1180954

[3] Yung-Ju Chang and John C Tang. 2015. Investigating mobile users' ringer mode usage and attentiveness and responsiveness to communication. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 6–15.

[4] Hee-Kyung Cho, Matthias Trier, and Eunhee Kim. 2017. The Use of Instant Messaging in Working Relationship Development: a Case Study. *Journal of Computer-Mediated Communication* 10, 4 (07 2017). https://doi.org/10.1111/j.1083-6101.2005.tb00280.x JCMC1044.

[5] 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* (New Orleans, Louisiana, USA) *(CSCW '02)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/587078.587081

[6] Trupti M Kodinariya and Prashant R Makwana. 2013. Review on determining number of Cluster in K-Means Clustering. *International Journal* 1, 6 (2013), 90–95.

[7] Hao-Ping Lee, Kuan-Yin Chen, Chih-Heng Lin, and Yung-Ju Chang. 2019. Connecting IM Pattern and Selective Perceived Responsiveness to Relationship: A Cluster-Based Approach. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1070–1074. https://doi.org/10.1145/3341162.3344841

[8] Hao-Ping Lee, Kuan-Yin Chen, Chih-Heng Lin, Chia-Yu Chen, Yu-Lin Chung, Yung-Ju Chang, and Chien-Ru Sun. 2019. *Does <i>Who</i> Matter? Studying the Impact of Relationship Characteristics on Receptivity to Mobile IM Messages*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300756

[9] Jun-Ki Min, Jason Wiese, Jason I. Hong, and John Zimmerman. 2013. Mining Smartphone Data to Classify Life-Facets of Social Relationships. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 285–294. https://doi.org/10.1145/2441776.2441810

[10] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.

[11] Jonathan Pettigrew. 2009. Text Messaging and Connectedness Within Close Interpersonal Relationships. *Marriage & Family Review* 45, 6-8 (2009), 697–716. https://doi.org/10.1080/01494920903224269 arXiv:https://doi.org/10.1080/01494920903224269

[12] 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* (Linz, Austria) *(MUM '15)*. Association for Computing Machinery, New York, NY, USA, 75–83. https://doi.org/10.1145/2836041.2836048

[13] Tainyi (Ted) Luor, Ling ling Wu, Hsi-Peng Lu, and Yu-Hui Tao. 2010. The effect of emoticons in simplex and complex task-oriented communication: An empirical study of instant messaging. *Computers in Human Behavior* 26, 5 (2010), 889–895. https://doi.org/10.1016/j.chb.2010.02.003 Advancing Educational Research on Computer-supported Collaborative Learning (CSCL) through the use of gStudy CSCL Tools.

[14] Peter Vorderer, Dorothée Hefner, Leonard Reinecke, and Christoph Klimmt. 2017. *Permanently online, permanently connected: Living and communicating in a POPC world.* Routledge.