Characterizing User Behavior on a Mobile SMS-Based Chat Service

Rafael de A. Oliveira, Wladmir C. Brandão & Humberto T. Marques-Neto

Instituto de Ciências Exatas e Informática
Pontifícia Universidade Católica de Minas Gerais (PUC)
Belo Horizonte – MG – Brazil
Agenda

1. Context
2. Overview
3. User behavior
4. Conclusions
5. Looking forward
Context

• Use of mobile instant messaging (IM) services has grown significantly last years

• 600M adults are currently using IM services on their mobile devices [Mander 2014]

• Understand the behavior of their users can help the improvement of user experience, performance, availability, cost, and quality of offered service.
Context

Dataset

• Mobile SMS-based chat service
• Major cellphone carrier from Brazil
• Data collected from May 10th to May 16th, 2014
• 2,348,805 messages
• 21,210 users

Each record contains:

• Session ID
• Sender
• Category
• Message
• Message Type
• Timestamp
**Context**

**Types of message**

- **Public Messages**
  - Visible to all

- **Room Messages**
  - Visible to all but directed

- **Private messages**
  - Sent to a single user (one-to-one)
Context

Room categorization

• General: messages of sports or religions
• Location: messages related to cities and regions
• Person: messages in personal chat rooms
• Relationship: messages about nightlife or flirting
Overview

• There is no rich interface like WhatsApp, Viber, etc
• SMS-Based

To talk to someone:

T + destination nickname + text message

• The service has an average of 335,000 messages per day (May 2014)
Overview

• Messages exchanging by day and by category

The highest amount of messages exchanged in a day occurs on Wednesday, corresponding to 14.95% of all exchanged messages in the week.
Overview

• Message exchanging throughout the day

As this service creates opportunities to entertainment and social relationships, we believe the evening massive usage is related to a kind of “social need” of users. The non-occurrence of a weekly fluctuation and the high use of service in the evenings could be explained by this need.
Overview

- User sessions by message type

> 87% of the user sessions we have exclusively Public and Room messages, suggesting a non-confidentiality pattern in the message exchanging.

Almost half of user sessions are exclusively formed by Room messages, which suggests that users mostly communicate pairwise, but without worrying about the privacy of the communication.
Overview

Messages by type on user sessions

77% of the messages are exchanged in non-confidential user sessions. This “open communication” suggests user interest for new relationships.

Many users build new relationships in non-confidential user sessions, and some of them intensify existing relationships in private user sessions.
User Behavior

• User Message Exchanging Distribution

The user message exchanging behavior follows a heavy-tailed distribution [Clauset et al. 2009], with a very small number of users sending the majority of the messages and the most of the users sending a very small number of messages on the chat service.
User Behavior

1) Define a feature set
   • Amount of
     • Messages
     • User sessions
   • Frequency
     • Message creation by minute

2) Apply a Clustering Algorithm
   • X-Means as the Clustering Algorithm [Pelleg et al. 2000]
     • Provides not only the clusters, but also estimating the suitable number of clusters
User Behavior

3) Cluster identification

• Light
  • A low frequency of use

• Infrequent
  • Moderate use
  • 6x/week

• Frequent
  • 20x/week.

• Heavy
  • Too many messages
  • 40x/week
### User Behavior – Weekly perspective

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Users</th>
<th>Messages</th>
<th>Sessions</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>Avg</td>
<td>CV</td>
<td>Avg</td>
</tr>
<tr>
<td>Light</td>
<td>65.00</td>
<td>33.16</td>
<td>1.59</td>
<td>1.55</td>
</tr>
<tr>
<td>Infrequent</td>
<td>25.00</td>
<td>156.08</td>
<td>0.94</td>
<td>6.26</td>
</tr>
<tr>
<td>Frequent</td>
<td>8.00</td>
<td>440.59</td>
<td>0.86</td>
<td>16.62</td>
</tr>
<tr>
<td>Heavy</td>
<td>2.00</td>
<td>934.63</td>
<td>0.99</td>
<td>36.47</td>
</tr>
</tbody>
</table>
User Behavior – Daily perspective

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Messages</th>
<th></th>
<th>Sessions</th>
<th></th>
<th>Frequency</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg</td>
<td>CV</td>
<td>Avg</td>
<td>CV</td>
<td>Avg</td>
<td>CV</td>
</tr>
<tr>
<td>Light</td>
<td>17.56</td>
<td>0.34</td>
<td>1.33</td>
<td>0.24</td>
<td>0.82</td>
<td>0.28</td>
</tr>
<tr>
<td>Infrequent</td>
<td>49.28</td>
<td>0.40</td>
<td>2.38</td>
<td>0.44</td>
<td>0.58</td>
<td>0.06</td>
</tr>
<tr>
<td>Frequent</td>
<td>112.41</td>
<td>0.39</td>
<td>4.41</td>
<td>0.55</td>
<td>0.60</td>
<td>0.11</td>
</tr>
<tr>
<td>Heavy</td>
<td>181.18</td>
<td>0.34</td>
<td>5.55</td>
<td>0.27</td>
<td>0.62</td>
<td>0.08</td>
</tr>
</tbody>
</table>
User Behavior

Looking backward

• 55% of heavy users from a specific day were classified into a different user profile on the previous day

Looking forward

• 85% of heavy users from a specific day return to the service on the day after

Heavy Users has a behavior of intensively exploiting service resources.
Behavioral changes

User Behavior
User Behavior

Categories exploitation
Conclusions

• Comprehensive characterization of the user behavior on a mobile SMS-based chat service provided by a major cellphone company in Brazil

• Different from previous work in literature, we provide a characterization of a private SMS-based chat service to detect malicious or atypical user behavior.

• Usage patterns of this service
  • Light users, Infrequent users, Frequent users and Heavy users

• Heavy Users tend to keep their behavior over time

• Heavy Users look for Relationship chat rooms
Looking forward

• Are Heavy Users a potential malicious users?
• Can we detect malicious behavior, such as defamation, pedophilia, phishing, and spamming from the message content?
• Are there another features to find better clusters?
• Are private messages being used for phishing?
Thank you!

Rafael de A. Oliveira, Wladmir C. Brandão & Humberto T. Marques-Neto