Exploring Timelines of Confirmed Suicide Incidents through Social Media

Xiaolei Huang, Linzi Xing, Jed R. Brubaker, Michael J. Paul
Departments of Information Science & Computer Science, University of Colorado Boulder

Abstract
Suicide is one of leading causes of death worldwide, yet little data is available about the lives of suicide victims because most people do not seek treatment[1]. Research has shown that people express suicidal ideation on social media[2], which can potentially be tapped to improve our understanding of the thoughts and behaviors of people prior to suicide. In this work, we introduce a novel dataset of Chinese social media accounts of 130 people who committed suicide between 2011&2016. We describe the demographic and geographic composition of the users, then conduct a longitudinal text analysis of their post histories, showing observable changes in content leading up to the time of death. With encouraging exploratory findings, we discuss directions for future research.

Background & Motivation
- 800,000 death[1]: Over 800,000 people die due to suicide per year, which is underestimated.
- 20 attempts[2]: For each adult who died of suicide, there may have been more than 20 others attempting suicide.
- 2nd leading cause of death[3,4]: Suicide is the 2nd leading cause of death in 15-29-year-olds.
- 7% of treatment coverage[4]: Chinese official statistics can't provide fine-grained information of suicide. Suicide is under-treated in China: only 7% of people who committed suicide had received health treatment.
- 300 million users[5]: In 2016, Weibo has over 300 million monthly active users, 82% of whom are 15 to 29 years old. Weibo might be a supplementary source to learn about suicidal issues in China.

Experiments & Approaches
- Time Series Analysis: We calculated statistics to summarize the activity levels and content of the users’ social media feeds in time intervals with respect to the date of death -10 months, 10 weeks, 10 days prior to death.
- Bootstrap resampling: We performed bootstrap resampling to measure uncertainty in our estimates in each time period. We created 100 estimates of the statistics, where each estimate was calculated by randomly sampling with replacement from the 130 users.

Measurements of Text (see Figure 3)

<table>
<thead>
<tr>
<th>Posting frequency</th>
<th>Linguistic indicators</th>
<th>Topic composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count-based</td>
<td>Lexicon-based</td>
<td>Topic Model-based</td>
</tr>
</tbody>
</table>

Analysis & Discussion
- Posting frequency: An increase in posting frequency in the last week before death, with a fairly steady increase through the final days.
- Linguistic indicators: It shows an increase in levels of language suggestive of suicidal ideation in the weeks leading up to death. There is high variability in the scores in the daily plot, so it is less clear what the trend is in the final days, but the score reaches high points in the 3–5 days before death.
- Topic composition: The topic distribution is stable over time, suggesting that users largely share similar content leading up to their suicide. But, we observe an increase in negative sentiment topics close to the time of death, with a decrease in the entertainment and daily life topics.
- Comparing Suicide Contexts: Broken relationship VS. Other Causes of Suicide. (see Figure 3 and 4)

PRIVACY AND ETHICS
All of the data has been anonymized. We only include aggregated information in this paper, and do not publish information associated with individual accounts (including example messages). We do, however, provide a description of how the data was collected and from where, so that the methodology can be replicated. Our approach is to make data available on a case-by-case basis upon request, sharing subsets of the data needed to only fulfill a research request.

References
2. Jared Jashinsky, Scott H Burton, Carl L Hanson, Josh West, Christophe Giraud-Carrier, Michael D Barnes, and Trenton Argyle. Tracking suicide risk factors through Twitter in the US. Crisis, 2014