CS 6474/CS 4803 Social Computing: Credibility I

Munmun De Choudhury
munmund@gatech.edu
Week 10 | October 23, 2017
Midterm Presentations/Milestone Reports

- 9 groups in all
- Presentations: Oct 30
- Reports due: Nov 1
- Structure of presentations same as the reports
- Each presentation will be about 7 minutes including the talk and Q&A
  - 5 minutes of talk
  - 2 minutes of Q&A
Assignment II Discussion
Reddit Apologizes For Speculating About Boston Marathon Suspects

Speculation on social media last week over who was responsible for the bombing at the Boston Marathon produced its own set of innocent victims: the falsely accused.

Reddit -- which was fiercely criticized for its "Findthebostonbombers" thread that called out specific people standing near the scene as suspects -- has now issued a public apology to those whose names were dragged through the mud.

"We all need to look at what happened and make sure that in the future we do everything we can to help and not hinder crisis situations," the statement reads.

The message, which was published in a blog post on Monday, continues by mentioning the serious ramifications of the rampant speculation that occurred on the site.
Information Credibility on Twitter
Summary (1)

- The paper investigates use of machine learning methods to assess credibility of tweets – distinguish news from rumor or fake information.
- Features used: message, user, topic, propagation.
- Findings:
  - User based features perform better than message features
  - Presence of URL in a post was the feature with highest information gain followed by estimate of negative sentiment and question-centric nature of posts.
- Human assessment to create ground truth for the algorithm. 86% accuracy.
- Limitations:
  - Only focused on tweets belonging to bursty topics.
Summary (2)

- Event identification from Twitter Monitor
- Separate news from personal opinion (Mechanical Turk)
- Credibility assessment of news (Mechanical Turk)
Summary (3)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVG REG AGE</td>
<td>1</td>
<td>1326</td>
<td>346</td>
<td>156</td>
</tr>
<tr>
<td>AVG STAT CNT</td>
<td>173</td>
<td>53841</td>
<td>6771</td>
<td>6627</td>
</tr>
<tr>
<td>AVG CNT FOLLOWERS</td>
<td>5</td>
<td>9425</td>
<td>842</td>
<td>946</td>
</tr>
<tr>
<td>AVG CNT FRIENDS</td>
<td>0</td>
<td>1430</td>
<td>479</td>
<td>332</td>
</tr>
<tr>
<td>FR HAS URL</td>
<td>0</td>
<td>1</td>
<td>0.616</td>
<td>0.221</td>
</tr>
<tr>
<td>AVG SENT SCORE</td>
<td>-2</td>
<td>1.75</td>
<td>0.038</td>
<td>0.656</td>
</tr>
<tr>
<td>FR SENT POS</td>
<td>0</td>
<td>1</td>
<td>0.312</td>
<td>0.317</td>
</tr>
<tr>
<td>FR SENT NEG</td>
<td>0</td>
<td>1</td>
<td>0.307</td>
<td>0.347</td>
</tr>
<tr>
<td>CNT DIST SHORT URLS</td>
<td>0</td>
<td>4031</td>
<td>121</td>
<td>419</td>
</tr>
<tr>
<td>SHR MOST FREQ AU</td>
<td>0</td>
<td>1</td>
<td>0.161</td>
<td>0.238</td>
</tr>
<tr>
<td>FR TW USER MENTION</td>
<td>0</td>
<td>1</td>
<td>0.225</td>
<td>0.214</td>
</tr>
<tr>
<td>FR TW QUEST MARK</td>
<td>0</td>
<td>1</td>
<td>0.091</td>
<td>0.146</td>
</tr>
<tr>
<td>FR EMOT SMILE</td>
<td>0</td>
<td>0.25</td>
<td>0.012</td>
<td>0.028</td>
</tr>
<tr>
<td>FR PRON FIRST</td>
<td>0</td>
<td>1</td>
<td>0.176</td>
<td>0.211</td>
</tr>
<tr>
<td>MAX LEV SIZE</td>
<td>0</td>
<td>632</td>
<td>46</td>
<td>114</td>
</tr>
</tbody>
</table>

Table 3: Summary for classification of newsworthy topics.

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (“true”)</td>
<td>0.825</td>
<td>0.108</td>
<td>0.874</td>
<td>0.825</td>
<td>0.849</td>
</tr>
<tr>
<td>B (“false”)</td>
<td>0.892</td>
<td>0.175</td>
<td>0.849</td>
<td>0.892</td>
<td>0.87</td>
</tr>
<tr>
<td>W. Avg.</td>
<td>0.860</td>
<td>0.143</td>
<td>0.861</td>
<td>0.860</td>
<td>0.86</td>
</tr>
</tbody>
</table>
Other examples

Controversy and Sentiment in Online News

Yelena Mejova*, Amy X. Zhang†, Nicholas Diakopoulos‡, Carlos Castillo*  
*Qatar Computing Research Institute, †University of Maryland, ‡MIT CSAIL  
ymejova@qf.org.qa, axz@mit.edu, nicholas.diakopoulos@gmail.com, chato@acm.org

ABSTRACT
How do news sources tackle controversial issues? In this work, we take a data-driven approach to understand how controversy interplays with emotional expression and biased language in the news. We begin by introducing a new dataset of controversial and non-controversial terms collected using crowd-sourcing. Then, focusing on 15 major U.S. news outlets, we compare millions of articles discussing controversial and non-controversial issues over a span of 7 months. We find that in general, when it comes to controversial issues, the use of negative affect and biased language is prevalent, while the use of strong emotion is tempered. We also observe many differences across news sources. Using these findings, we show that we can indicate to what extent an issue is controversial, by comparing it with other issues in terms of how they are portrayed across different media.

Our contribution. In this work, we quantify the use of emotional and biased language when presenting controversial issues in the news. We begin by building a list of controversial and non-controversial terms in current news in the U.S. using crowd-sourcing techniques. Then, we perform a large-scale analysis of millions of news articles from 15 U.S.-based news sources. We focus on the expression of sentiment using a series of lexical resources containing words conveying positive and negative emotions; this automatic analysis helps reduce the inherent subjectivity of traditional content analysis methods.

We demonstrate that controversial issues in news can be characterized by the use of fewer positive words and a greater presence of negative words. This finding is consistent across different media sources and confirmed with 4 different sentiment lexicons. Interestingly, we find that the use of highly emotional terms (as opposed to mild ones) is less likely in the context of controversial topics, suggesting a self-moderation on the part of the news sources.
The paper does not exploit the wealth of information embedded in the network structure of a user. This can be very useful for credibility purposes. Discuss some ways you’d utilize this information.
The authors stop short of proposing any sort of user-specific model (although they use user-features): beyond attributes of the information and the information sharing agent, it is necessary to model end user perceptions.

For instance, some users may value sentiment more than social connections, which may in turn impact how they assess information credibility.
Are non-experts (e.g., Turkers) the right people to assess credibility?
A need for “fact checking systems” that operate outside of the social media ecosystem. But these systems are difficult to build and use. Why?
Tweeting is Believing? Understanding Microblog Credibility Perceptions
Summary

• The article presents results of a survey study that gauged users’ perceptions of tweet credibility.
• 26 features identified to impact credibility assessment based on a pilot sample of five individuals.
• Two controlled experiments to measure the impact of several tweet features (message topic, user name, and user image) on perceptions of message and author credibility
  • Participants recruited from inside Microsoft and through a message board of CMU
• Total of 256 completed surveys, 101 from the corporate group and 155 from the alumni group.
• Main findings:
  • Users are poor judges of tweet credibility based on content alone; reliance on cues such as username and presence/absence of profile pictures.
Newsworthiness and Network Gatekeeping on Twitter: The Role of Social Deviance

Nicholas Diakopoulos
Columbia University, School of Journalism
nicholas.diakopoulos@gmail.com

Arkaitz Zubiaga
Dublin Institute of Tech., Applied Intelligence Research Centre
arkaitz@zubiaga.org

Abstract
Publishers of news information are keen to amplify the reach of their content by making it as re-sharable as possible on social media. In this work we study the relationship between the concept of social deviance and the re-sharing of news headlines by network gatekeepers on Twitter. Do network gatekeepers have the same predilection for selecting socially deviant news items as professionals? Through a study of 8,000 news items across 8 major news outlets in the U.S. we predominately find that network gatekeepers re-share news items more often when they reference socially deviant events. At the same time we find and discuss exceptions for two outlets, suggesting a more complex picture where newsworthiness for networked gatekeepers may be moderated by other effects such as topicality or varying motivations and relationships with their audience.

network gatekeepers. Previous literature has found that events with high deviance were more likely to gain coverage in the main stream media (MSM) (P. Shoemaker, Danielian, and Brendlinger 1991), “the role of news media is not to mirror the world as it is, but rather to spotlight and draw public attention to problems and situations that need solutions and repair,” (Pamela Shoemaker 2006). Shoemaker’s theory posits that MSM will select for and favor socially deviant stories and events.

But do network gatekeepers share a similar predilection for selecting socially deviant news items? Is social deviance a professionally imbued newsworthiness criterion, or something that more generally explains interest, attention, and sharing of news? Networked gatekeeping theory explores the idea that every user on Twitter is a gatekeeper, with the discretion to share or not
One limitation of the work is that their current recruitment method does not include certain demographics that consume tweets, like teenagers or adults without a college degree; education may matter.

The paper focused on a rather well-educated and specialized group of participants, and that it failed to contrast results of this population and a more general population.
Class Reading – Significance of assessing credibility of anti-vax information
Both papers focus on assessing credibility of news. Would same observations apply to judging credibility of non-real time information? E.g., health myths
Credibility is, after all, a domain-dependent attribute. What additional new feature would you consider, in addition to the ones raised in the papers, that could be useful for this purpose? How would you factor in end users’ bias in perception of credibility? (Take the example of the anti-vaxx health myth)
Which of these pictures of Hurricane Sandy are Real and which ones are fake?
7 Fake Hurricane Sandy Photos You're Sharing on Social Media
Class Exercise II

The papers examined and studied credibility on Twitter – a primarily text based content system. How would the different cues change if we look at the host of new multimedia sharing social apps (e.g., Instagram) and want to assess credibility of such content?
Modulating Video Credibility via Visualization of Quality Evaluations

Nicholas Diakopoulos
School of Communication and Information
Rutgers University
4 Huntington St., New Brunswick, NJ, USA
diakop@rutgers.edu

Irfan Essa
School of Interactive Computing
Georgia Institute of Technology
75 5th St. NW, Atlanta, GA, USA
irfan@cc.gatech.edu

ABSTRACT
In this work we develop and evaluate a method for the syndication and visualization of aggregate quality evaluations of informational video. We enable the sharing of knowledge between motivated media watchdogs and a wider population of casual users. We do this by developing simple visual cues which indicate aggregated activity levels and polarity of quality evaluations (i.e. positive / negative) which are presented in-line with videos as they play. In an experiment we show the potential of these visuals to engender constructive changes to the credibility of informational video under some circumstances. We discuss the limitations, and future work associated with this approach toward video credibility modulation.

information quality by combing through the media and engaging in fact-checking and re-contextualization of news and other media reports. For high profile video events such as the State of the Union address given by the president of the U.S., there is a considerable demand for this type of watchdogging activity. For instance, recent coverage by news outlets like PBS included annotated transcripts and video snippets showing analysis from experts and journalists. One of the major issues with such analytic presentations as are found on Politifact, Factcheck, and PBS is that, especially for video, the analysis is divorced from the video itself, making the multimedia context difficult to understand in relationship with the textual analysis.

While most methods of watchdogging are labor intensive, another