2017

Deception Detection & Rumor Debunking

Victoria L. Rubin

Western University, vrubin@uwo.ca

Follow this and additional works at: https://ir.lib.uwo.ca/fimspres

Part of the Library and Information Science Commons

Citation of this paper:
https://ir.lib.uwo.ca/fimspres/45
Deception Detection & Rumor Debunking

VICTORIA RUBIN

ASSOCIATE PROFESSOR,
Faculty of Information & Media Studies (FIMS)

DIRECTOR,
Language & Information Technology Research Lab (LiT.RL)

Western University
Faculty of Information & Media Studies

LiT.RL
Language & Information Technology Research Lab

Victoria Rubin. CCA2017. Congress of Humanities and Social Sciences, Ryerson, Toronto, 05/30/2017
Goals of the Talk

I’ll divide my 10-minute talk into 2 parts: (1) Deception Detection and (2) Rumor Debunking, as the title suggests, and I will argue for the need of hybrid methods (in a combination of the two).

My main goal here is to point researchers interested in social media research towards these 2 exciting fields.

I predict that such technologies (with more R&D, as they mature) will become indispensable in our attention-economy.

Content producers are rushed to be first in the news stream, and social media consumers simply don’t have time or energy to verify content that is pushed at them.
Part 1.
Deception

a message knowingly and intentionally transmitted to foster a false belief or conclusion

Buller & Burgoon (1996) in Communication Theory
Zhou et al. (2004) in Group Decision and Negotiation

an intentional control of information in a technologically mediated environment

Hancock (2012) in Oxford Handbook of Internet Psychology
Detection

Human Ability To Detect Deception

55–58% success rate
Frank et al. 2004; Kraut, 1980; Vrij, 2000

54% mean accuracy
DePaulo et al., 1997

Social Psychology & Communications Studies
Detection

Recently proven possible

at ~74% accuracy with Natural Language Processing

Zhou et al., 2004

at ~70% with Machine Learning

Mihalcea & Strapparava, 2009
Verbal **Cues** for Automated Deception Detection

Deceivers:
- self-references
- detailed answers
- indirect statements
- negative emotions

**PATTERNS IN PREDICTORS EXIST, BUT NO CLEAR CONSENSUS. VARIATIONS BY CONTEXTS.**

Hancock et al. (2008) in *Discourse Processes*

Larker & Zakolyukina (2010) featured in *The Economist*

Are there cues of deception in social media?

Photo by Eric Pickersgill, www.removed.social/, The NY Review of Books
‘Butler Lies’ in Texting

manage or avoid social interactions

(‘Yeah sorry I gtg’)

“(Hancock, et al., 2009; French et al, 2015)”
Part 2.
Rumors

Unverified assertions... spread over time from node to node in a network.
Vosoughi (2015), MIT PhD Thesis

Harmful. Why?

Undesirable responses: defamation, protests, destruction of properties, spread of fear or hate, euphoria, or stock market fluctuations.
Matthews (2013), Time

Photo by Ben White, unsplash.com
Figure 3. Verification Feature for Rumor Debunking on Twitter (Liu et al., 2015).

The six proposed categories of verification features largely based on insights from journalists.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>FEATURE NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOURCE CREDIBILITY</td>
<td>Is trusted/satirical news account</td>
</tr>
<tr>
<td></td>
<td>Has trusted/satirical news url</td>
</tr>
<tr>
<td></td>
<td>Profile has url from top domains</td>
</tr>
<tr>
<td></td>
<td>Client application name</td>
</tr>
<tr>
<td>SOURCE IDENTITY</td>
<td>Profile has person name</td>
</tr>
<tr>
<td></td>
<td>Profile has location</td>
</tr>
<tr>
<td></td>
<td>Profile includes profession information</td>
</tr>
<tr>
<td>SOURCE DIVERSITY</td>
<td>Has multiple news/non-news urls after dedup</td>
</tr>
<tr>
<td></td>
<td>Deduped tweets’ text is dissimilar</td>
</tr>
<tr>
<td>SOURCE LOCATION &amp; WITNESS</td>
<td>If tweet location matches event location</td>
</tr>
<tr>
<td></td>
<td>If profile location matches event location</td>
</tr>
<tr>
<td></td>
<td>Has witness phrases, i.e., “I see” and “I hear”</td>
</tr>
<tr>
<td>MSG. BELIEF</td>
<td>Is support, negation, question or neutrality</td>
</tr>
<tr>
<td>EVENT PROPAGATION</td>
<td>Event Topic</td>
</tr>
<tr>
<td></td>
<td>Retweet, mention, hashtag h-index</td>
</tr>
<tr>
<td></td>
<td>Max reply/retweet graph4 size/depth</td>
</tr>
</tbody>
</table>

(Liu et al, 2015, Reuters)
Final Thoughts...

✓ Hybrid approaches are needed

✓ More R&D needed based on social media data.

✓ Detailed R&D overview (Chapter 21).

✓ Come to my talk on the News Verification Suite @ CAIS Wed May 31 @ 2.
Thank you! Questions? Ideas?

Contact – vrubin@uwo.ca
http://victoriarubin.fims.uwo.ca/
References upon request and in the book chapter.