2017

Deception Detection & Rumor Debunking

Victoria L. Rubin
Western University, vrubin@uwo.ca

Deception Detection & Rumor Debunking

VICTORIA RUBIN

ASSOCIATE PROFESSOR,
Faculty of Information & Media Studies (FIMS)
DIRECTOR,
Language & Information Technology Research Lab (LiT.RL)

Western University · Canada
Western FIMS
Faculty of Information & Media Studies

LiT RL
Language & information Technology Research Lab
VICTORIA RUBIN

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

Photo by Nicole Mason, unsplash.com

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

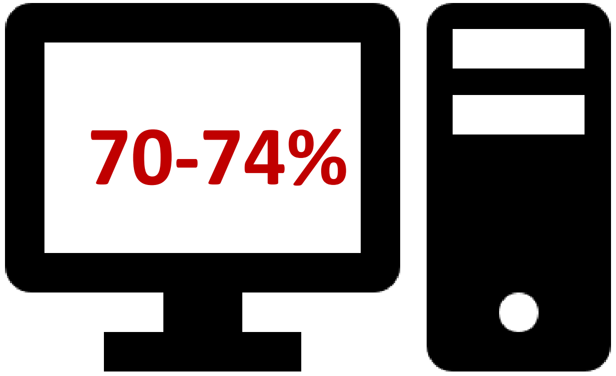
at ~74% accuracy with Natural Language Processing

Zhou et al., 2004

at ~70% with Machine Learning

Mihalcea & Strapparava, 2009

Automated Deception Detection



Verbal **Cues** for Automated Deception Detection

Deceivers:

self-references

detailed answers

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

indirect statements

negative emotions

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

sense-based word

Granhag et al. (2004) in *Legal & Criminological Psych.*

Are there cues of deception

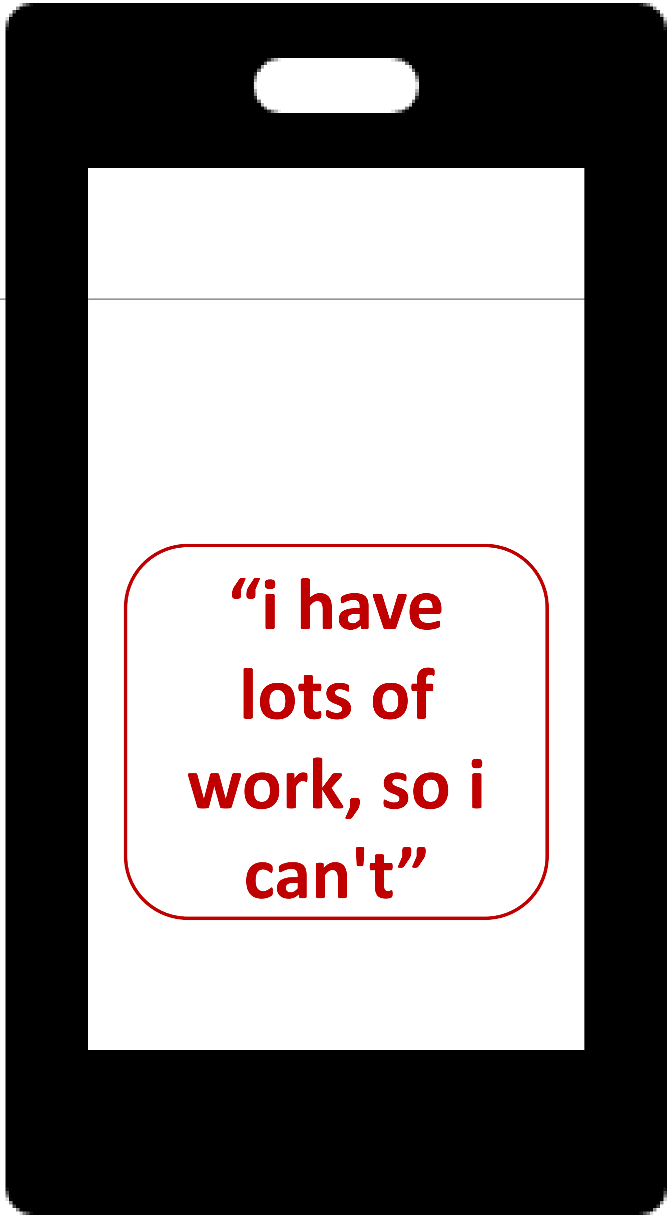
in social media?

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”



“i have lots of work, so i can't”

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.

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

(Liu et al, 2015, Reuters)

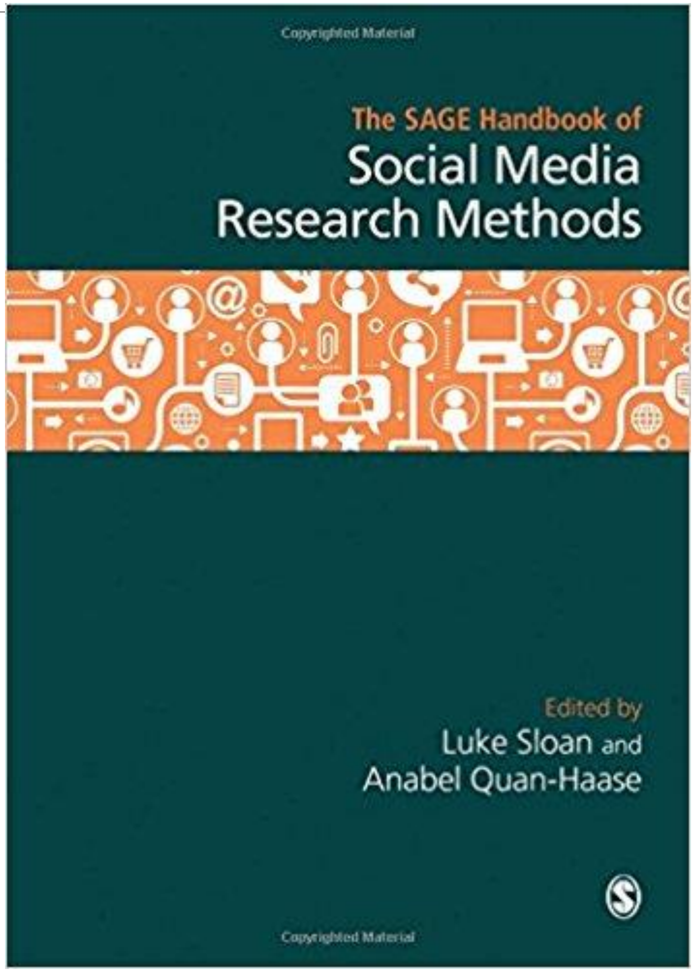
Final Thoughts…

✓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.

Western
UNIVERSITY · CANADA

FIMS
Faculty of Information & Media Studies

LiT RL Language & information Technology Research Lab

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada