Recommender Systems and Misinformation:

The problem or the solution?

Miriam Fernández, Open University, United Kingdom
Alejandro Bellogín, Universidad Autónoma de Madrid, Spain

Examples of misinformation

Word of the Year 2016

After much discussion, debate, and research, the Oxford Dictionaries Word of the Year 2016 is... *post-truth*.

Post-truth is an adjective defined as 'relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief'.

Why was this chosen?

The concept of *post-truth* has been in existence for the past decade, but Oxford Dictionaries has seen a spike in frequency this year in the context of the EU referendum in the United Kingdom and the presidential election in the United States. It has also become associated with a particular noun, in the phrase *post-truth politics*.

The term has moved from being relatively new to being widely understood in the course of a year – demonstrating its impact on the national and international consciousness. The concept of *post-truth* has been simmering for the past decade, but Oxford shows the word spiking in frequency this year in the context of the Brexit referendum in the UK and the presidential election in the US, and becoming associated overwhelmingly with a particular noun, in the phrase *post-truth politics*.

The Economist ✔
@TheEconomist

Obama founded ISIS. George Bush was behind 9/11. Welcome to post-truth politics

Art of the lie
Politicians have always lied. Does it matter if they l entirely?
🔗 economist.com

12:28 PM · Nov 1, 2016

♡ 307 💬 387 people are Tweeting about this

The Independent ✔
@Independent

'We've entered a post-truth world and there's no going back'

We've entered a post-truth world and there's no going back
How did we come to a mass state of altered consciousness, as foreseen by George Orwell? And how did we come to it so quickly?
🔗 independent.co.uk

4:10 PM · Nov 8, 2016

30 💬 22 people are Tweeting about this

https://languages.oup.com/word-of-the-year/2016/

Examples of misinformation

Coronavirus: The human cost of virus misinformation

By Marianna Spring
Specialist disinformation and social media reporter

🕐 27 May

Coronavirus pandemic

GETTY IMAGES

A BBC team tracking coronavirus misinformation has found links to assaults, arsons and deaths. And experts say the potential for indirect harm caused by rumours, conspiracy theories and bad health information could be much bigger.

"We thought the government was using it to distract us," says Brian Lee Hitchens, "or it was to do with 5G. So we didn't follow the rules or seek help sooner."

Brian, 46, is talking by phone from his hospital bed in Florida. His wife is critically ill - sedated, on a ventilator in an adjacent ward.

"The battle that they've been having is with her lungs," he says, voice wobbling. "They're inflamed. Her body just is not responding."

After reading online conspiracy theories, they thought the disease was a hoax - or, at the very least, no worse than flu. But then in early May, the couple caught Covid-19.

"And now I realise that coronavirus is definitely not fake," he says, running out of breath. "It's out there and it's spreading."

https://www.bbc.com/news/stories-52731624

3

Misinformation spread

Factors that influence the spread of misinformation online:

- How <u>information</u> is constructed and presented

- Personality, values, and emotions of the <u>users</u>

- Characteristics of <u>digital platforms</u>

- **<u>Algorithms</u> that power the recommendation of information in those platforms**

Our goal

Explore the effect of existing algorithms on the recommendation of false and misleading information
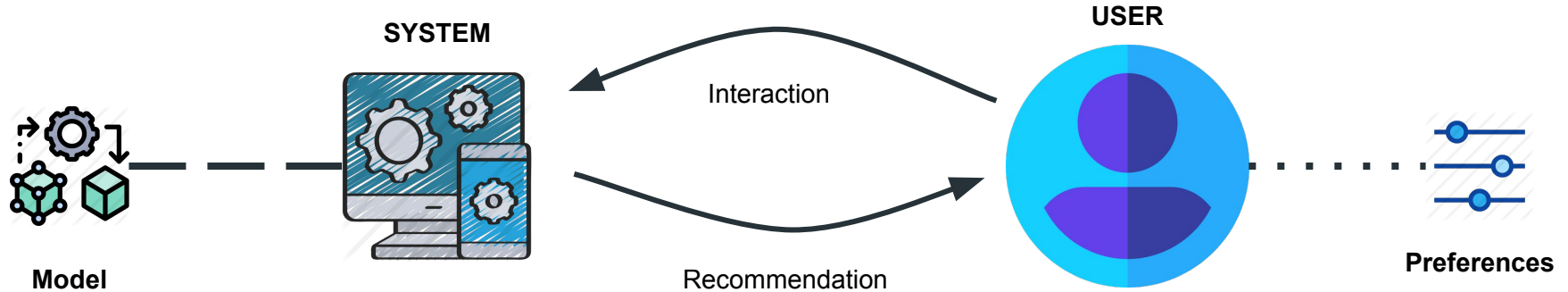
Understand which techniques are more prone to spread misinformation, and under which circumstances.

How can their internal functioning be modified, or adapted, to counter this behaviour?

Current solutions

Facebook found a better way to fight fake news

KEY POINTS

- Facebook says the Related Articles feature gives people more context about the story, with the added benefit of reducing the frequency a hoax article is shared.

- After a year of testing, the company found that Disputed Flags inadvertently buried critical information that explained the inaccuracies, and could backfire by entrenching a person's false beliefs.

https://www.cnbc.com/2017/12/21/facebook-found-a-better-way-to-fight-fake-news.html

How Twitter's algorithm is amplifying extreme political rhetoric

New York (CNN Business) — Imagine opening up the Twitter app on your phone and scrolling through your feed. Suddenly, you come across a hyper-partisan tweet calling Hillary Clinton the "godmother of ISIS." It's from a user you do not follow, and it's not in your feed by virtue of a retweet from a user you do follow. So how did it get there?

Over the last several months, Twitter has begun inserting what it believes to be relevant and popular tweets into the feeds of people who do not subscribe to the accounts that posted them. In other words, Twitter has started showing users tweets from accounts that are followed by those they follow. This practice is different from the promoted content paid for by advertisers, as Twitter is putting these posts into the feeds of users without being paid and without consent from users.

https://edition.cnn.com/2019/03/22/tech/twitter-algorithm-political-rhetoric/index.html

Current "solutions"

Facebook found a better way to fight fake news

PUBLISHED THU, DEC 21 2017·9:01 AM EST   UPDATED THU, DEC 21 2017·9:01 AM EST

THE VERGE | Thuy Ong

SHARE f 𝕏 in ✉

KEY POINTS

- Facebook says the Related Articles feature gives people more context about the story, with the added benefit of reducing the frequency a hoax article is shared.

- After a year of testing, the company found that Disputed Flags inadvertently buried critical information that explained the inaccuracies, and could backfire by entrenching a person's false beliefs.

https://www.cnbc.com/2017/12/21/facebook-found-a-better-way-to-fight-fake-news.html

How Twitter's algorithm is amplifying extreme political rhetoric

By Oliver Darcy, CNN Business

Updated 1142 GMT (1942 HKT) March 22, 2019

New York (CNN Business) — Imagine opening up the Twitter app on your phone and scrolling through your feed. Suddenly, you come across a hyper-partisan tweet calling Hillary Clinton the "godmother of ISIS." It's from a user you do not follow, and it's not in your feed by virtue of a retweet from a user you do follow. So how did it get there?

Over the last several months, Twitter has begun inserting what it believes to be relevant and popular tweets into the feeds of people who do not subscribe to the accounts that posted them. In other words, Twitter has started showing users tweets from accounts that are followed by those they follow. This practice is different from the promoted content paid for by advertisers, as Twitter is putting these posts into the feeds of users without being paid and without consent from users.

https://edition.cnn.com/2019/03/22/tech/twitter-algorithm-political-rhetoric/index.html
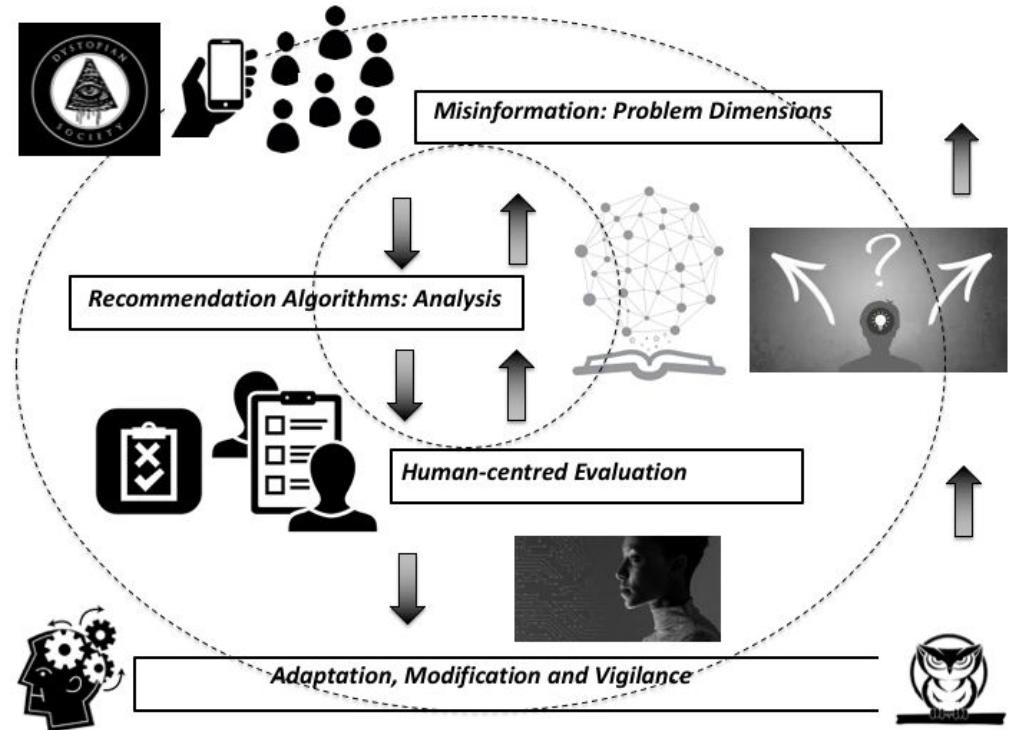
Our approach

which recommendation algorithms are more prone to spread misinformation, and under which circumstances?

Problem Dimensions: which dimensions of the misinformation problem affect the behaviour of RS?

Analysis of Recommendation Algorithms: analyse how existing algorithms improve or worsen the spread of misinformation

Human-centred Evaluation: modify existing evaluation methods and metrics (dissonance vs. satisfaction)

Adaptation, Modification and Vigilance: adapt existing algorithms and track their impact over time (vigilance).

Misinformation: Problem Dimensions

Recommendation Algorithms: Analysis

Human-centred Evaluation

Adaptation, Modification and Vigilance

8

Misinformation: Problem Dimensions

Dimensions that intersect between recommendation and misinformation

Content: Different *forms* (newspaper articles, blog posts, social media posts), *topics* (health, elections*), formats* (text, images, videos), *framing* (false news, rumours, conspiracy theories), *origin* (news outlets, social contacts, public figures), *time*

Users: Different *motivations*, *personalities*, *values*, *emotions* and *susceptibility* influence the spread of misinformation

Platform & Network Features: Platforms are designed differently (*content limitations, sharing permissions*) & have different *typology and topology of network structure.*

Others: *global events, presence of malicious actors, presence of checked facts*

A global problem,
multiple dimensions
(human, sociological, technological)

# Analysis of Recommendation Algorithms

Which recommendation techniques are more prone to suggest misinformative items to users and why?

CF methods neglect any information about the items and their misinformative features. They can help us to understand the impact of user-item interactions.

CB approaches use content and therefore require analysing NL and its inherent subtleties.

Hybrid methods combine content analysis with user-time interactions.
Are they therefore more prone to spread misinformation?

User & Item Profiles: Multiple representations can be considered.
How do these representations influence the spread of misinformation?
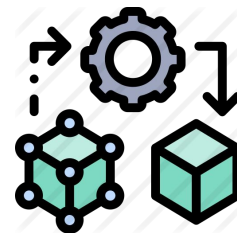
Human-centred Evaluation

Metrics that focus on computing a balanced degree of user satisfaction and discomfort (introducing opposing views) may help to combat misperceptions.
Same information, opposing views ≠ diversity (novel information)

Datasets: We need users / items & information about their credibility + user-item ratings

- NELA-GT-2018 dataset (713 misinforming articles) → no user profiles / ratings
- COVID-19 Poynter dataset (603 misinforming articles/posts) → no user profiles / ratings
- NewsReel: users, articles and ratings → no information about item credibility

Measure: how can we evaluate when the task has been successfully addressed?

Adaptation, Modification, and Vigilance

Social Science and Psychology studies: simply presenting people with corrective information is likely to fail in changing their salient beliefs.

- Providing an explanation rather than a simple refute
- Exposing users to related but disconfirming stories
- Revealing similarities with the opposing groups

Adaptation/Modification: the adaptation of existing algorithms to counter the misinformation recommendation problem should build on existing social science and psychology theory.

- Profile users to better capture motivations
- RS that introduce small degrees of opposing views
- RS focus on similarity and dissimilarity
- Explanations

Vigilance:

- Dynamics of misinformation captured over time
- Algorithms are adapted
- Proposed adaptations do not back-fire (ethics!)

Discussion

Recommender systems are often accused of feeding and reinforcing the interaction cycle

BUT, they could be part of the **solution**, if we understand:

- How misinformation is spread
- How such mechanisms are reinforced
    - And to what extent, depending on the technique
- How we can adapt the algorithms to counter these effects

Our hypothesis: build on <u>social science</u> and <u>psychology</u> theories

Current plan

Starting with a COVID-related Poynter dataset
    It contains news items classified as reliable or not

We are collecting Twitter user profiles
    Challenge 1: how many users can we collect for each news item? *(cold items)*
    Challenge 2: how many items do user profiles actually contain? *(cold users)*
    Challenge 3: what is a *realistic* ratio between misinformation/normal items?

    Question: are there datasets with this information combined?

Next step: run and evaluate CF recommenders
    Next challenges???

Fact-Checked by: TheJournal.ie
2020/09/18 | Ireland
FALSE: The swab test for COVID-19 can enter the brain and cause infections

Read More

Fact-Checked by: Chequeado
2020/09/17 | Argentina
MISLEADING: A chain that recommends how to avoid a relapse of the new coronavirus disease

Read More

Fact-Checked by: FactCheck.org
2020/09/16 | United States
FALSE: "Coronavirus was created in Wuhan lab and released intentionally."

Read More

Fact-Checked by: Chequeado
2020/09/16 | Argentina
FALSE: An interview with Conicet's senior researcher Hugo Luján, with the title: "Argentina already has the oral COVID-19 vaccine and it is not being taken into account"

Read More

Recommender Systems and Misinformation:

The problem or the solution?

Miriam Fernández, Open University, United Kingdom
Alejandro Bellogín, Universidad Autónoma de Madrid, Spain