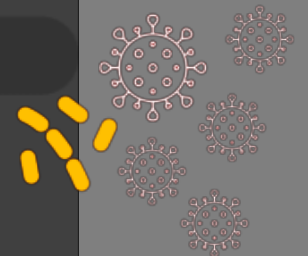


Deception in Medical Information across Natural Language Genres

 Ritwik Banerjee

 rbanerjee@cs.stonybrook.edu

Computer Science, Stony Brook University



**Integrating Conflicting Data:
The Role of Source Dependence**

Corroborating Information from Disagreeing Views*

Xin Luna
AT&T Labs-F
Florham Park,
lunadong@res

The Wisdom of Minority: Discovering and
Targeting the Right Group of Workers for
Crowdsourcing

**DeClarE: Debunking Fake News and False Claims
using Evidence-Aware Deep Learning**

Kashyap Popat¹, Subhabrata Mukherjee², Andrew Yates¹, Gerhard Weikum¹

¹Max Planck Institute for Informatics, Saarbrücken, Germany

²Amazon Inc., Seattle, USA

Credibility : [EMNLP 2018](#)

The wisdom of minority: [WWW 2014](#), [COLING 2014](#)

Truth discovery: [VLDB 2009](#), [WSDM 2010](#)

**Integrating Conflicting Data:
The Role of Source Dependence**

Corroborating Information from Disagreeing Views*

Xin Luna
AT&T Labs-F
Florham Park,
lunadong@res

The Wisdom of Minority: Discovering and
Targeting the Right Group of Workers for
Crowdsourcing

**DeClarE: Debunking Fake News and False Claims
using Evidence-Aware Deep Learning**

Kashyap Popat¹, Subhabrata Mukherjee², Andrew Yates¹, Gerhard Weikum¹

¹Max Planck Institute for Informatics, Saarbrücken, Germany

²Amazon Inc., Seattle, USA

**Cues to deception and ability to detect lies as a
function of police interview styles**

Aldert Vrij¹, Samantha Mann, Susanne Kristen, Ronald P Fisher

Credibility : [EMNLP 2018](#)

The wisdom of minority: [WWW 2014](#), [COLING 2014](#)

Truth discovery: [VLDB 2009](#), [WSDM 2010](#)

Deceptive language : [Law & Human Behav. 2007](#)

**Integrating Conflicting Data:
The Role of Source Dependence**

Xin Luna
AT&T Labs-F
Florham Park,
lunadong@res

Corroborating Information from Disagreeing Views*

The Wisdom of Minority: Discovering and
Targeting the Right Group of Workers for
Crowdsourcing

**DeClarE: Debunking Fake News and False Claims
using Evidence-Aware Deep Learning**

Kashyap Popat¹, Subhabrata Mukherjee², Andrew Yates¹, Gerhard Weikum¹
¹Max Planck Institute for Informatics, Saarbrücken, Germany
²Amazon Inc., Seattle, USA

**Cues to deception and ability to detect lies as a
function of police interview styles**

**Seeing through deception: A computational approach to deceit
detection in written communication**

Ángela
English D
University

**The Lie Detector: Explorations in the Automatic Recognition
of Deceptive Language**

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

Syntactic Stylometry for Deception Detection

Myl

Song Feng Ritwik Banerjee Yejin Choi
Department of Computer Science
Stony Brook University
Stony Brook, NY 11794-4400

Credibility : [EMNLP 2018](#)

The wisdom of minority: [WWW 2014](#), [COLING 2014](#)

Truth discovery: [VLDB 2009](#), [WSDM 2010](#)

Deceptive language: [ACL 2009 - 2012](#)

Deceptive language : [Law & Human Behav. 2007](#)

**Integrating Conflicting Data:
The Role of Source Dependence**

Xin Luna
AT&T Labs-F
Florham Park,
lunadong@res

Corroborating Information from Disagreeing Views*

The Wisdom of Minority: Discovering and
Targeting the Right
Crowdsourcing

**DeClarE: Debunking Fake
using Evidence-Awa**

Kashyap Papat¹, Subhabrata Mukherje
¹Max Planck Institute for Inform
²Amazon Inc.,

**Cues to deception and ability to detect lies as a
function of police interview styles**

**Seeing through deception: A computational approach to deceit
detection in written communication**

Ángela **The Lie Detector: Explorations in the Automatic Recognition
of Deceptive Language**

Stretch of the Imagination

tion Detection

Yejin Choi

ence

00



Did we solve
misinformation detection?

Credibility : [EMNLP](#)

The wisdom of minority: [WWW 2014](#), [COLING 2014](#)

Truth discovery: [VLDB 2009](#), [WSDM 2010](#)

Deceptive language: [ACL 2009 - 2012](#)

Deceptive language : [Law & Human Behav. 2007](#)



CNN

+ Follow

Covid misinformation is everywhere

By Brian Stelter and Virginia Langmaid, CNN Business - Nov 9

No, a 'Detox' Bath Will Not Reverse Your Covid-19 Vaccine

Anti-vaxxers have pivoted to "undoing" vaccines with epsom salt and borax, which will absolutely not work.

By Whitney Kimball | 11/12/21 7:15PM | Comments (124) | Alerts


How to Detox Vaccines and Antidote COVID-19 Inoculations (A Global Perspective)

May 14, 2021 / Dr. Gerald H. Smith / No Comments

HOW TO DETOXYFY AND HEAL FROM VACCINATIONS – FOR ADULTS AND CHILDREN

May 22, 2015 by Michael Edwards

Last updated on: August 28, 2021

 CNN [+ Follow](#)

Covid misinformation is everywhere

By Brian Stelter and Virginia Langmaid, CNN Business - Nov 9

No, a 'Detox' Bath Will Not Reverse Your Covid-19 Vaccine

Anti-vaxxers have pivoted to "undoing" vaccines with epsom salt and borax, which will absolutely not work.

By Whitney Kimball · 10/21 2:05PM · Comments (24) · Alerts

Blurred boundary between information and opinion

All we get are **claims!**

How to Detox Vaccines and Antidote COVID-19 Inoculations (A Global Perspective)

May 14, 2021 / Dr. Gerald H. Smith / No Comments

HOW TO DETOXYFY AND HEAL FROM VACCINATIONS – FOR ADULTS AND CHILDREN

May 22, 2015 by Michael Edwards

Last updated on: August 28, 2021

Detecting false claims

- Large datasets of labeled claims, rumors, etc.
- Labels obtained from fact-checking agencies or crowdsourcing
- Each claim labeled as true or false (binary classification) or somewhere in the true-false spectrum (regression)

Verified or not?
Rumor detection.

Verified to true or false?
Fact checking.

Worth verifying?
Check-worthiness.

For or against the claim?
Stance detection.



Why do we believe what we believe?

A prior belief system

- A healthy skepticism of our belief system: individual reality monitoring
- A healthy skepticism of external information: social/cultural reality monitoring

Verification

- But with respect to which sources?

Epistemological concerns

Ostermeier (2011), Uscinski (2015)

Verified or not?
Rumor detection.

Verified to true or false?
Fact checking.

Worth verifying?
Check-worthiness.

For or against the claim?
Stance detection.

Sources of verification

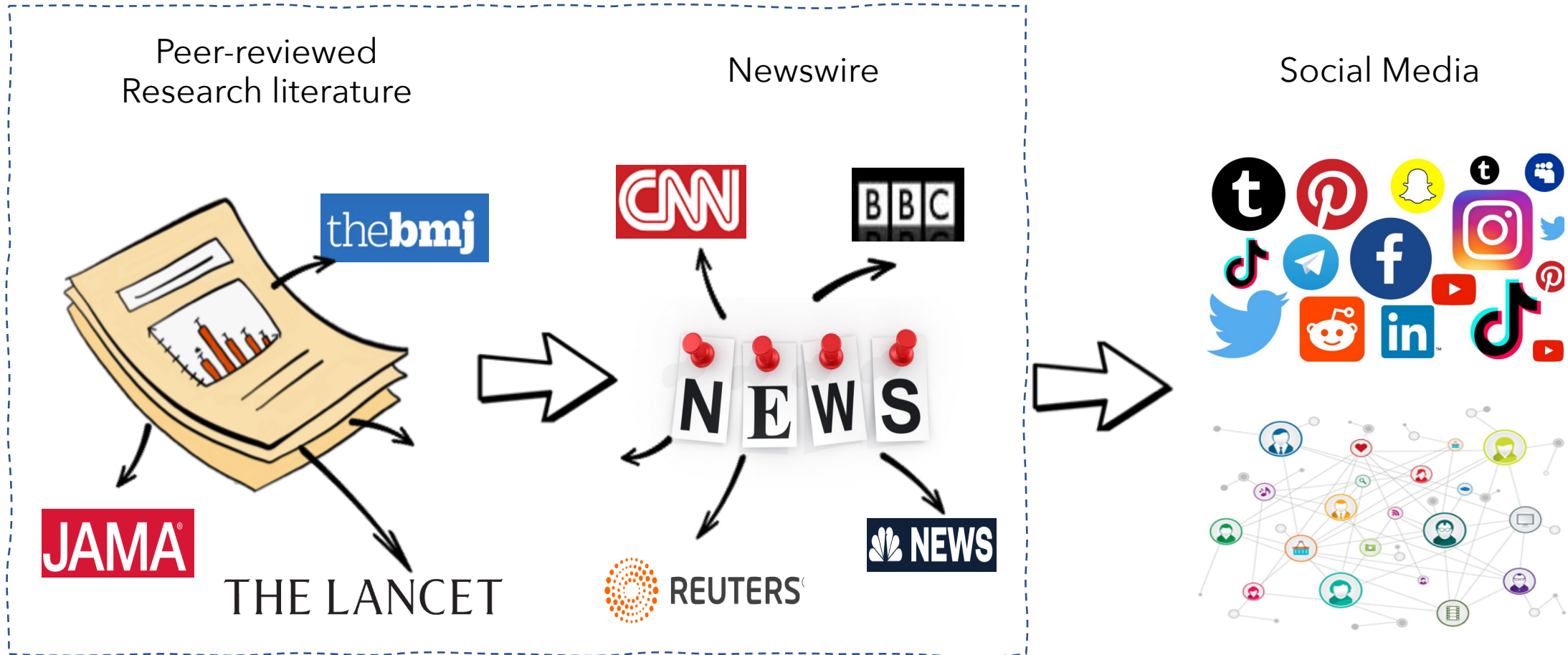
The source must be **credible**

- Authority of the authors
- Referrals to other credible sources

Rains (2009), Flanagin (2011), Corritore (2012)

What happens if credibility is a matter of perception?

Cross-genre information propagation



Cross-genre information propagation

Similarly, the **thyroid drug levothyroxine** should be taken “on an empty stomach, one-half to one hour before breakfast”.

<https://www.nytimes.com/2018/09/28/well/live/drug-medication-empty-stomach-prescription.html>

Comprehension of Top 200 Prescribed Drugs in the US as a Resource for Pharmacy Teaching, Training and Practice

[Andrea V. Fuentes](#), [Moises D. Pineda](#), and [Kalyan C. Nagulapalli Venkata](#)*

Surface-level contextual markers frequently determine the credibility of the text

Especially when readers

- have low topic-knowledge
- are unable to distinguish between nuanced claims made in technical language [Braten et al. \(2011\)](#)

What to verify?

In a new research published in the *Journal of the American Academy of Child and Adolescent Psychiatry*, researchers found that less parental warmth and having harsher home environments can contribute to how aggressive children become.

Flaxseed fiber reportedly helps balance cholesterol levels and lower blood pressure, among other benefits.

Now, a new study published in the *American Journal of Physiology: Endocrinology and Metabolism* suggests that flaxseed fiber can also lower obesity markers.

Health workers have also been using a new, experimental Ebola vaccine to help stop the virus' spread, inoculating almost 40,000 people so far. Health workers have been using a vaccine made Merck ^{???}, which has been **shown in field testing** to potentially reduce infection rates, and showed some success during the last outbreak in the DRC. Merck has provided thousands of the vaccines to health

Can it be verified?

- Upheld by empirical observations
- The observations are independent of subjective interpretation Ducasse (1935), Fleck (1975)

Bayesian confirmation!

- We cannot verify against every available source
- Outcome depends on *what* and *how much* was considered as evidence
- Evidence is pre-specified by the embedded hyperlink in the news article 😊

Data

- 229,000 health news articles
- 35,900 contain hyperlinks to peer-reviewed medical research



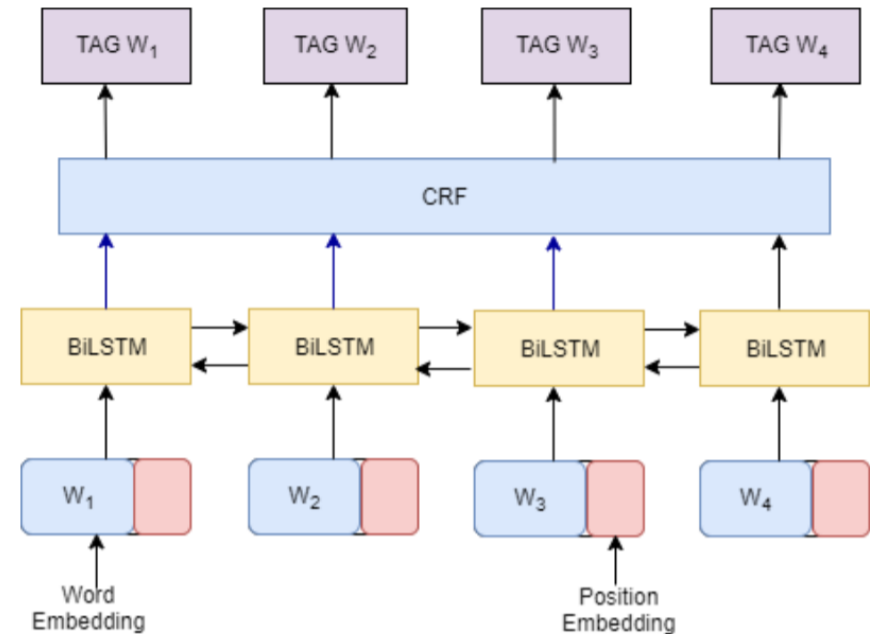
The experimental setup

Baseline

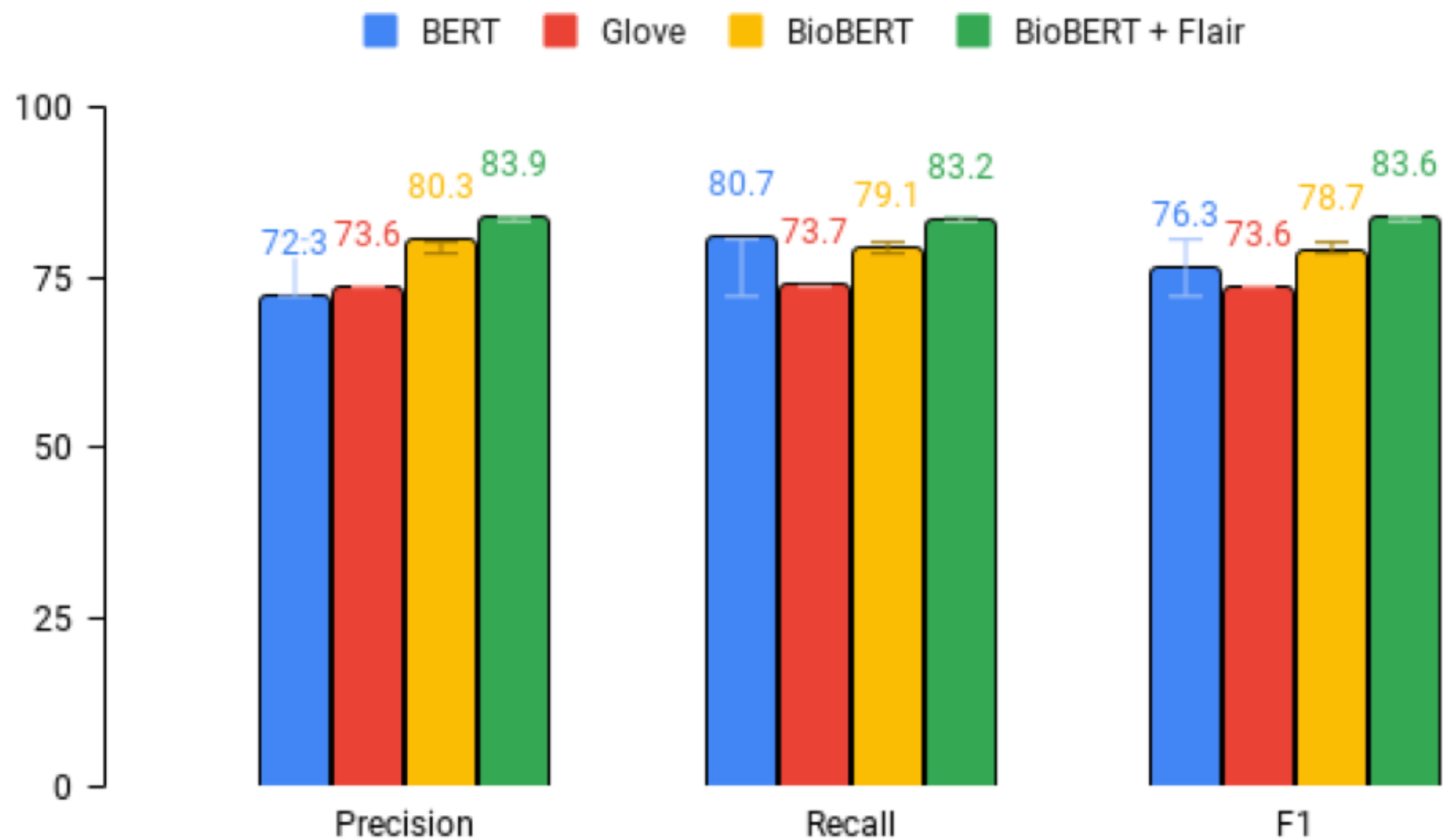
- BERT

+ BiLSTM-CRF

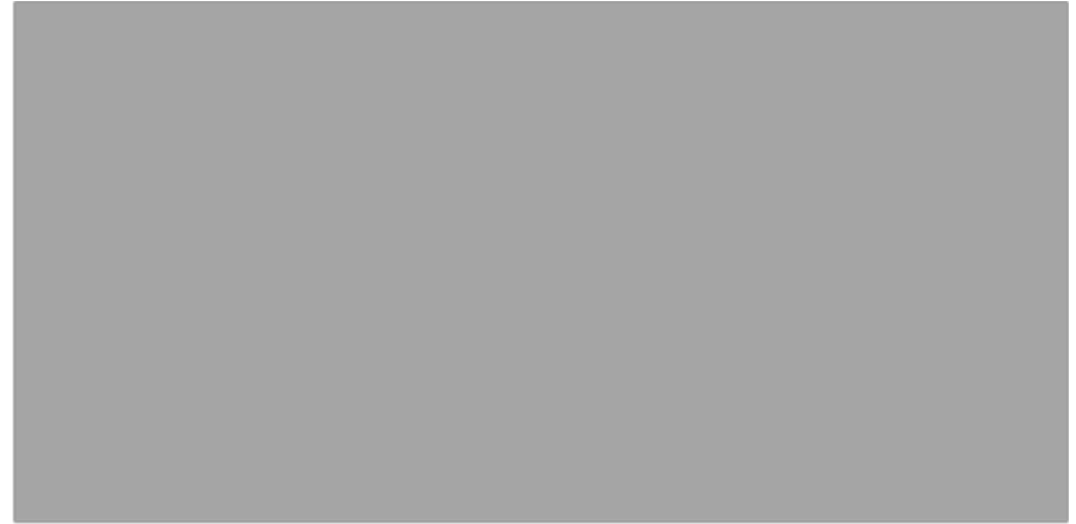
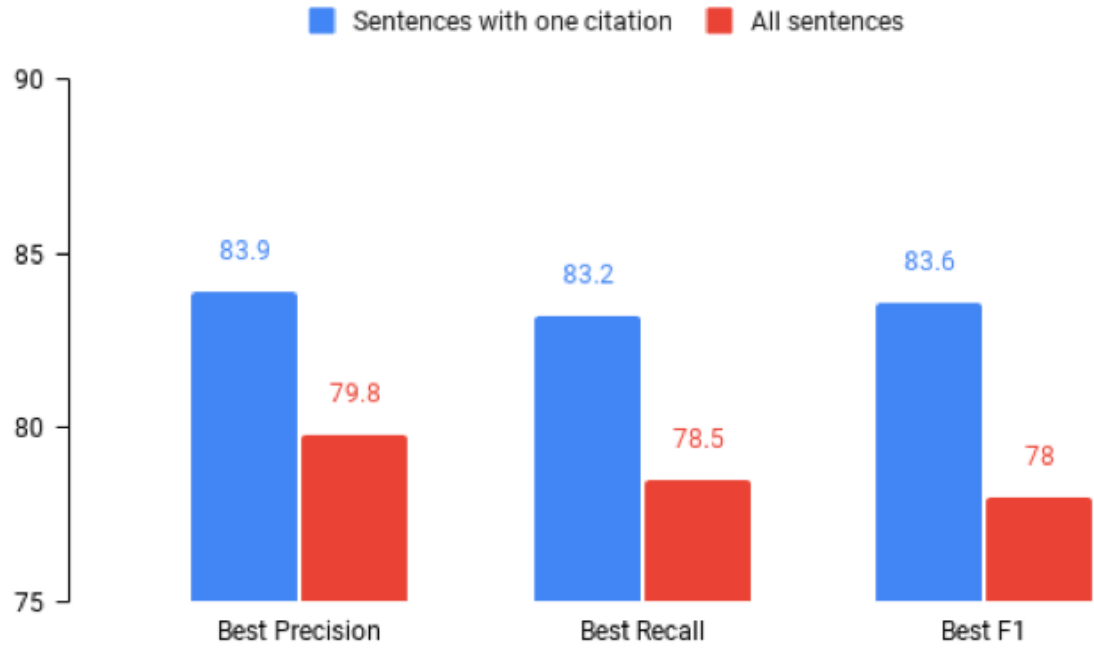
- Classical embeddings
- Contextual embeddings
- Stacked embeddings



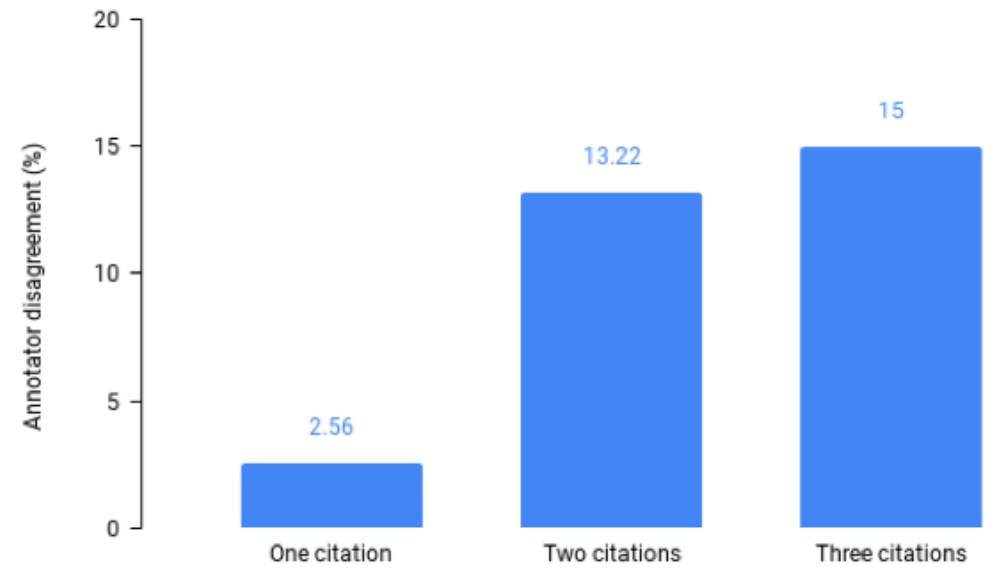
Model Comparison (on sentences with a single hyperlink)



Comparison across complexity of visual cues



Annotator disagreement (%)



Cross-genre verification



(claim, abstract)

(claim, sentence-1)

(claim, sentence-2)

(claim, sentence-3)

⋮

Pharmacy (Basel) 6(2), 2018 doi: 10.3390/pharmacy6020043



Pharmacy (Basel). 2018 Jun; 6(2): 43.

PMCID: PMC6025009

Published online 2018 May 14. doi: [10.3390/pharmacy6020043](https://doi.org/10.3390/pharmacy6020043)

PMID: [29757930](https://pubmed.ncbi.nlm.nih.gov/29757930/)

Comprehension of Top 200 Prescribed Drugs in the US as a Resource for Pharmacy Teaching, Training and Practice

[Andrea V. Fuentes](#), [Moises D. Pineda](#), and [Kalyan C. Nagulapalli Venkata](#)*

▶ [Author information](#) ▶ [Article notes](#) ▶ [Copyright and License information](#) ▶ [Disclaimer](#)

This article has been [cited by](#) other articles in PMC.

Abstract

Go to:

Pharmacists have access to a plethora of information related to drugs. Online compendia concerning top 200 prescribed drugs are readily-accessible, comparatively-easy to search. While these resources provide some information about the commonly prescribed drugs, they lack in furnishing in-depth knowledge to pharmacy students, pharmacists and other healthcare professionals. The aim of this paper is to present the relevant details of top 200 most prescribed drugs in the United States. The names and therapeutic classes of top 200 prescribed drugs were compiled from online resources. The pharmacological actions of drugs, any reported adverse reactions and black box warnings are collected from drug bank resources, such as AccessPharmacy and Lexicomp. The paper provides comprehensive information about top 200 prescribed drugs, which includes generic names, pharmacological action, route of administration and adverse reaction profile including black box warning when applicable. Overall, the drug list may serve as an easy access of ideas for pharmacists, researchers and other healthcare professionals interested in developing new strategies for treating patients with various ailments.

Keywords: black box warning, oral administration, cardiovascular, central nervous system, endocrine,

Cross-genre verification

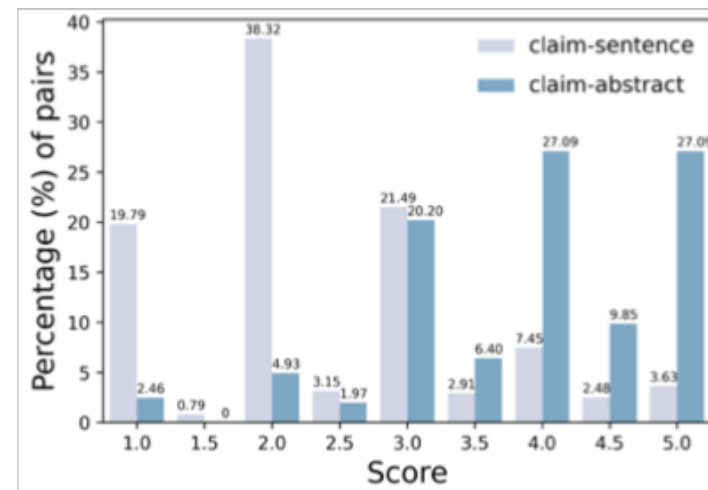
- Binary classification into *true* and *false* is a massive oversimplification of nuanced medical findings reported in research literature

[Ioaniddis \(2017\)](#), [Scheufele \(2019\)](#)

- Likert-like scale for claim-sentence pairs

1	s and c are completely unrelated, no inference is possible.
2	s does not describe the same event as c, but there are shared entities (usually, some relevant properties of those entities are being described).
3	s does not describe the same event as c, but c may still be inferred from s (typically based on expert domain-knowledge).
4	s contains some of the information in c, but some details are missing and may possibly be inferred (typically based on expert domain-knowledge).
5	s contains all the information in c, and thus, c can be immediately inferred from s.

Cross-genre verification

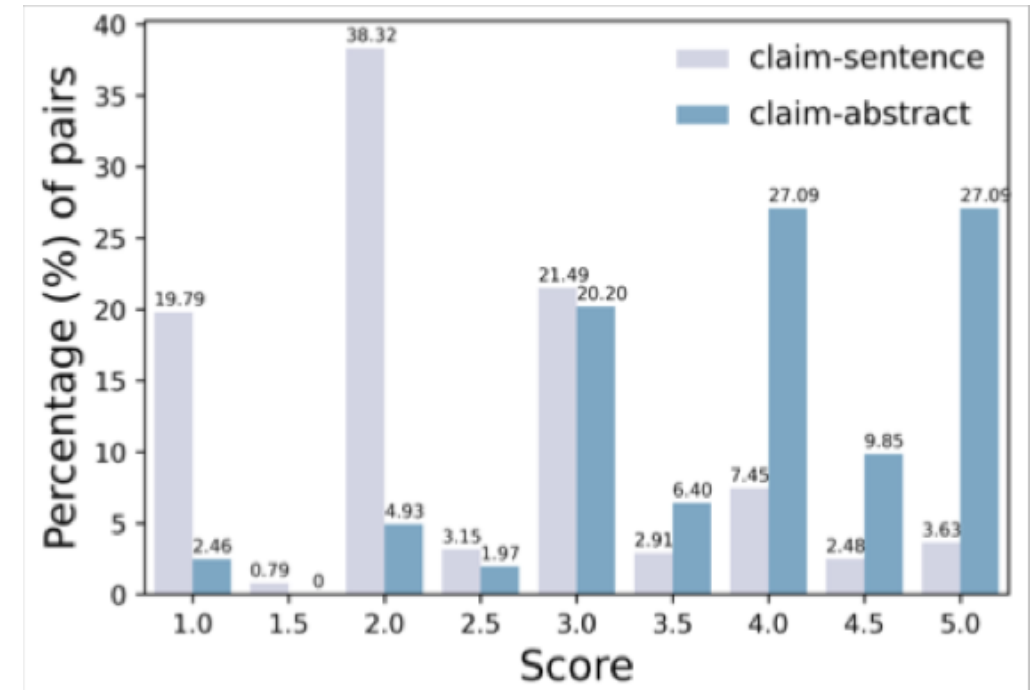


- Finally, we want to assign one of the three labels to each claim-abstract pair:
 - supported
 - unsupported
 - uncertain

$$\mathcal{L}(c, a) = \begin{cases} \text{supported} & \text{if } \max_{s \in a} \text{Likert}(c, s) = 5, \\ \text{unsupported} & \text{if } \max_{s \in a} \text{Likert}(c, s) \leq 2, \\ \text{uncertain} & \text{if } \max_{s \in a} 2 < \text{Likert}(c, s) \leq 4 \end{cases}$$

Insights from annotation

- 1,652 claim-sentence pairs (157 disagreements, only by 1)
- On the sentences where raters agree
 - positive skewness (0.44) for claim-sentence pairs
 - negative skewness (-0.27) for claim-abstract pairs



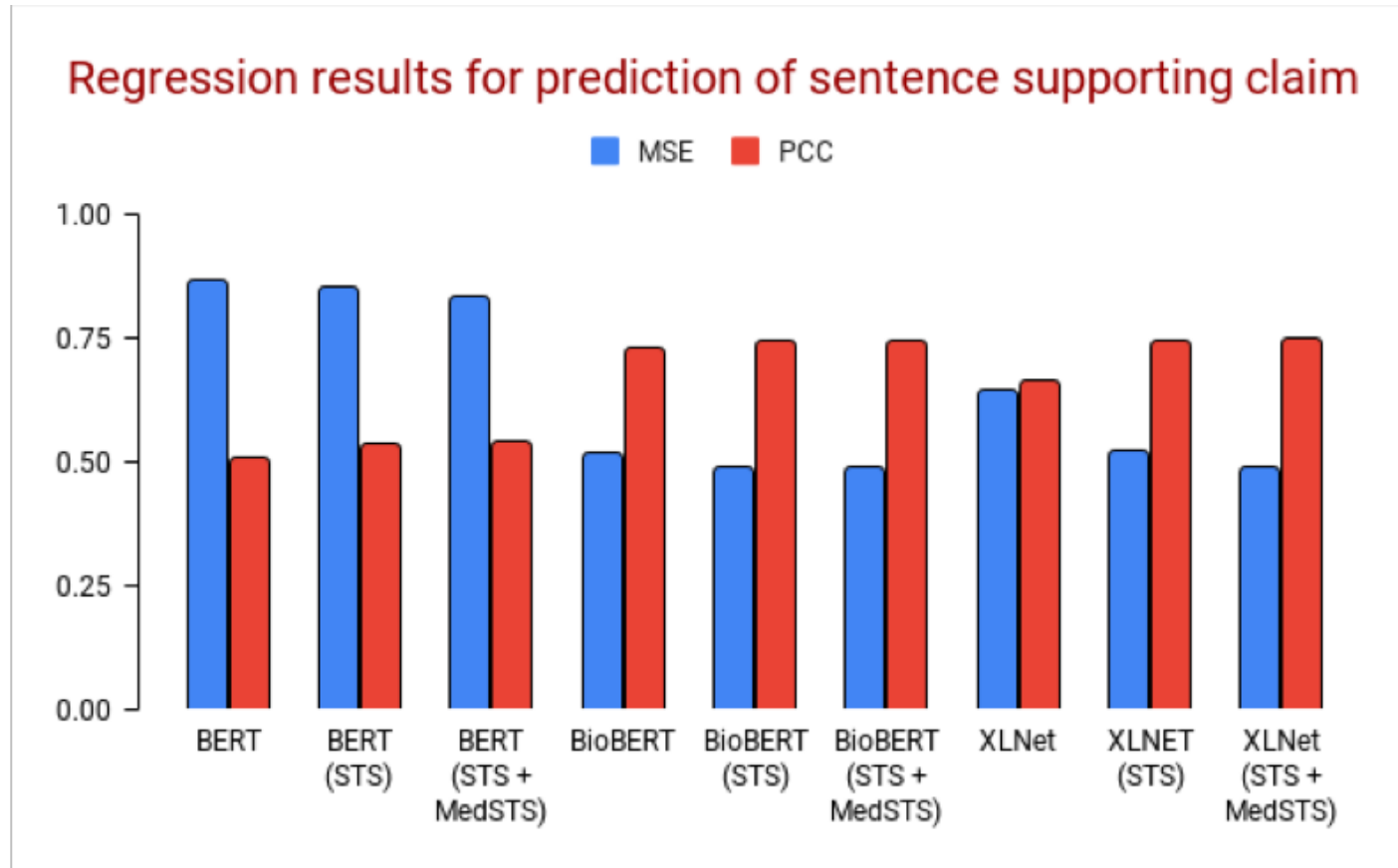
Support from a sentence

- Regression learning with target set [1, 5]
- Pretrained transformers: BERT, BioBERT, XLNet

Model tuning

- Task-based tuning on the semantic similarity benchmark (STS) [Cer et al. \(SemEval 2017\)](#)
- Domain-specific task-based tuning on the medical semantic similarity benchmark (MedSTS) [Wang et al. \(LREC 2018\)](#)

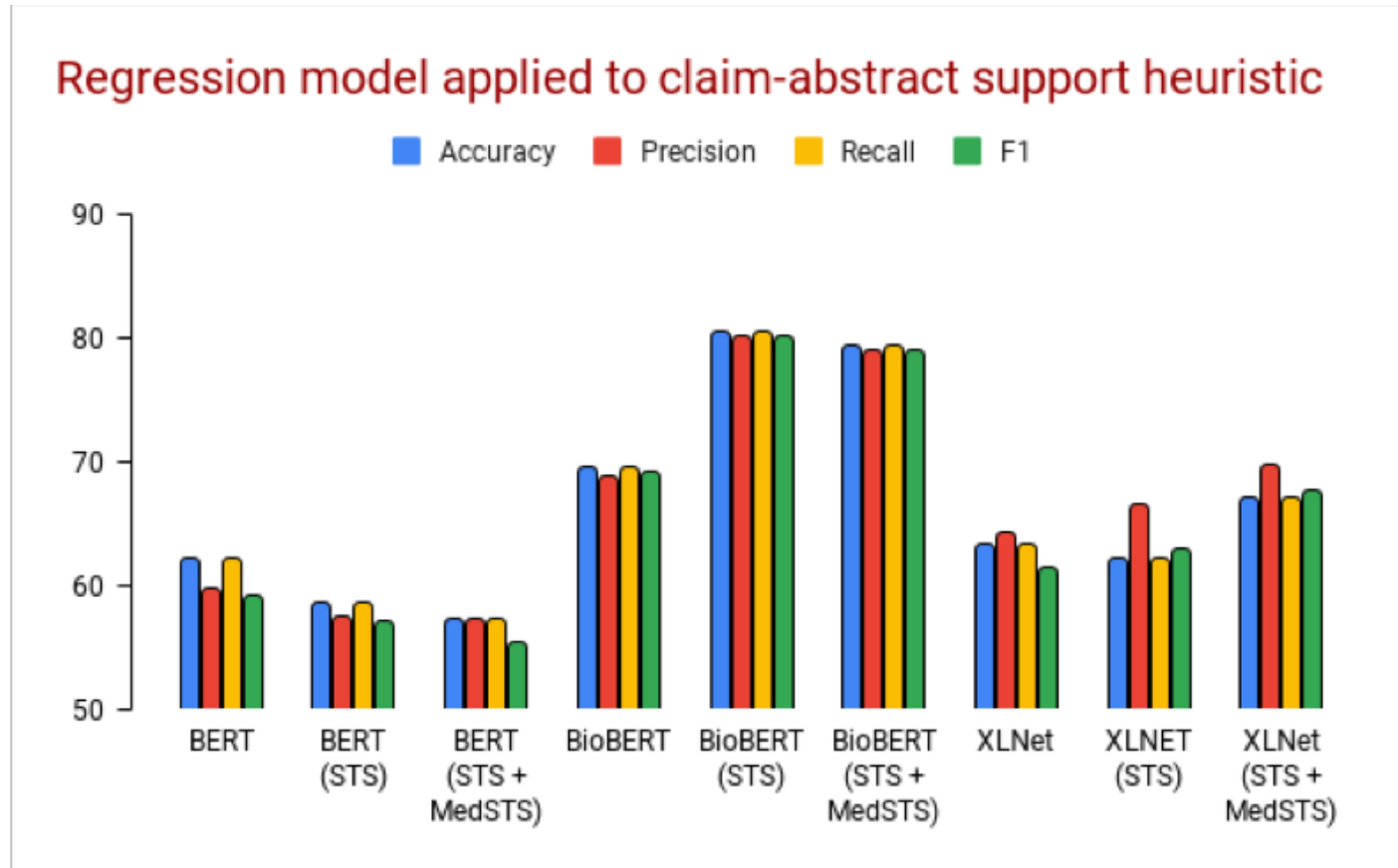
Support from a sentence



Evaluation

- Mean squared error (MSE)
- Pearson correlation coefficient (PCC)

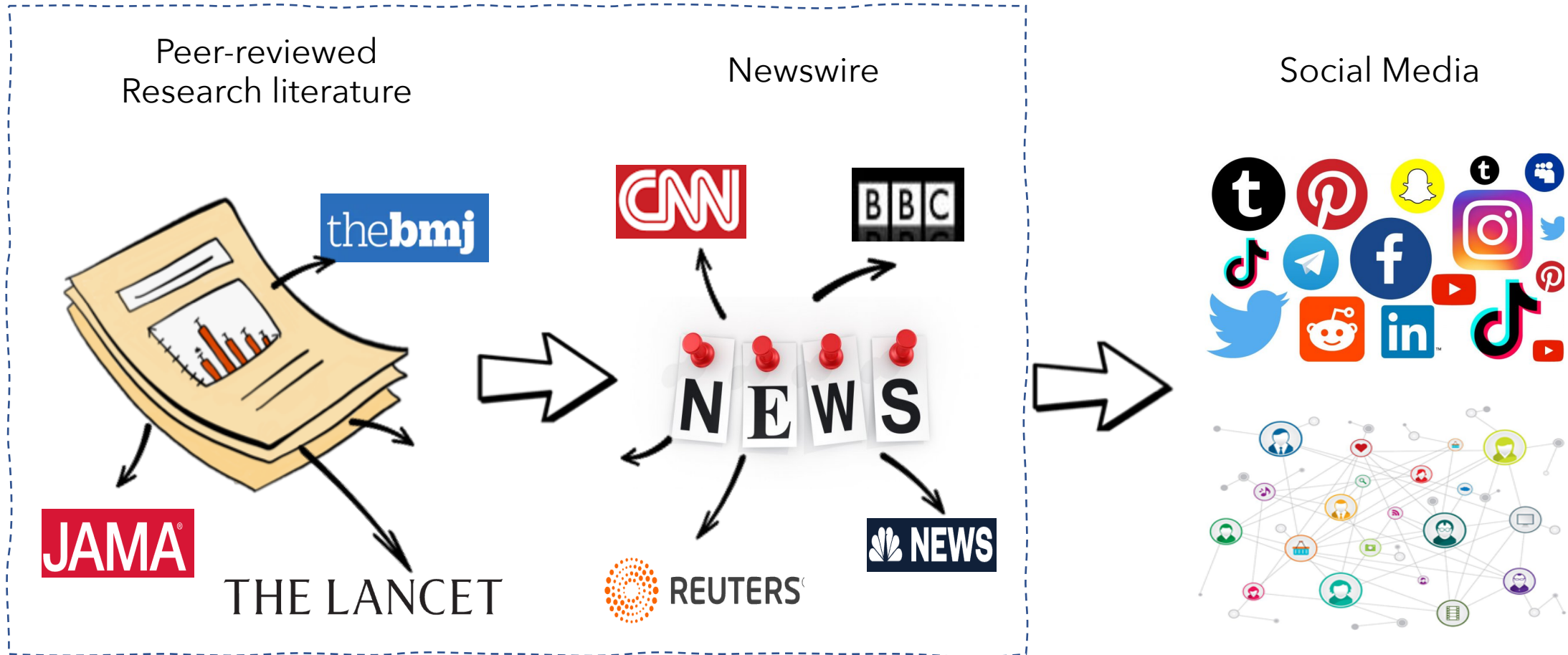
Support from the cited research abstract



Evaluation

- Plug-in the regression model's predictions into the heuristic function, rounding $f(x) = \frac{1}{2} [2x]$
- Micro-averaged accuracy
- Precision, recall, and F1

Cross-genre information propagation



What happens if credibility is a matter of perception?

We trust that perception!



Credibility
has been

taken at face value,
based on prior
belief/bias

built on
assumptions (e.g.,
credible articles
use objective
language)

- They may work for the “low hanging” fruit:
conspicuous propaganda sources
- Not so obvious for RT, Xinhua, or the New York Times
[Hersch \(2013\)](#)



1

We have barely scratched the surface: implication

Less-than-lying is more dangerous than lying Schauer and Zeckhauser (2009)

- Implied and indirect deception with natural language

🕒 Thursday Mar 26 2020



Kanika Kapoor met Prince Charles before he tested positive for COVID-19?

2

We have barely scratched the surface: distortion

(1)	Increase in the amount of exercise has been linked to lower risk of cardiovascular diseases in elderly men.
(2)	Increase in the amount of exercise has been linked to lower risk of cardiovascular diseases.
(3)	More exercise lowers risk of cardiovascular diseases.
(4)	More exercise reduces risk of diseases.

3

We have barely scratched the surface: whataboutism

- We maintain a healthy level of reality monitoring by verifying claims
- Verify with sources we consider credible
- What happens when, instead of countering the claim, an adversary attacks the credibility of the sources we rely on?

4

selection bias

5

whitewashing

6

slanting

Collaborators (in reverse chronological order)



- ★ Dr. Chaoyuan Zuo
- ★ Noushin Salek-Faramarzi
- ★ Khiem Phi
- ★ Kritik Mathur
- ★ Dhruv Kela



- ★ Dr. Indrakshi Ray
- ★ Dr. Hossein Shirazi
- ★ Fatemeh Hashemi



- ★ Dr. Hossein Shirazi



- ★ Dr. Yejin Choi



- ★ Dr. Song Feng



Q&A

