

AI, critica e qualitative analysis

Marco Dotti

marco.dotti@unipv.it

Bologna, 13 giugno 2024

Fake news detectors vs. debunker

I rilevatori di fake news si concentrano su autori e fonti. Ma cosa succede se la fonte è sconosciuta?

I rilevatori sono solitamente addestrati e testati sugli stessi argomenti di notizie e possono semplicemente catturare una distorsione nel set di dati a livello di argomento.

I modelli (patterns) di linguaggio

Una strada; analizzare **modelli di linguaggio** presenti nel corpo dell'articolo, **indipendentemente** dalla fonte o dall'autore.

È possibile creare un rilevatore di fake news che sfrutta **le reti neurali profonde** e, valutando un argomento che non è incluso nel set di dati di addestramento, può catturare modelli che non sono specifici dell'argomento.

Black box & Language war

Un problema ben noto del deep learning, ovvero **la mancanza di trasparenza nel processo decisionale**, noto anche come problema della *scatola nera*.

Il fatto di non sapere cosa sia alla base delle decisioni prese dalla rete lascia **l'incertezza di potersi fidare o meno dell'algoritmo**.

The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors

Nicole O'Brien^{1,3} Sophia Latessa^{1,3} Georgios Evangelopoulos^{1,3} Xavier Boix^{1,2,3}

¹Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

²Children's Hospital, Harvard Medical School, USA

³Center for Brains, Minds and Machines (CBMM)

{njobrien, sofiabl, gevang, xboix}@mit.edu

Questo studio rivela che la rete impara a **rilevare modelli linguistici** negli articoli di fake news **che possono essere generalizzati** per individuare fake news che coprono argomenti distinti.

Questo è un primo passo verso la comprensione dell'affidabilità dei rilevatori di fake news basati sul deep learning.

FAKE NEWS DETECTOR*

A deep learning network developed by CBMM computer scientists that detects patterns in the language of fake news.

Replace this text with your article. Be sure to remove ads and photos as this material will skew the results.

Analyze Language

Our team at the [Center for Brains, Minds and Machines](#) within MIT's [McGovern Institute for Brain Research](#) has developed a ["fake news detector"](#) that uses deep neural networks to capture subtle differences in the language of fake and real news.

For more information about this detector, [read this MIT news story](#) or take a look at our original manuscript, "[The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors](#)," which we presented at a workshop called "AI for Social Good" at the 32nd Conference on Neural Information Processing Systems (NIPS) in Montreal, Canada.

* Please note: This website is for research purposes only and should not be construed as a valid measure of the accuracy or legitimacy of any content submitted to this detector. [Read more](#)



This material is based upon work supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Come si addestra un rilevatore?

I ricercatori hanno addestrato una rete neurale con circa 24.000 articoli di notizie pubblicati tra il 26 ottobre e il 25 novembre 2016. Queste date sono state scelte appositamente perché abbracciano il periodo di tempo immediatamente precedente, durante e successivo alle elezioni presidenziali degli Stati Uniti del 2016.

Il set di dati sulle fake news consisteva in circa 12.000 articoli estratti da Kaggle, che conserva una lista nera di siti web di fonti di fake news.

Il set di dati sulle notizie reali comprende 9.000 articoli del Guardian e oltre 2.000 articoli del *New York Times*.

Cosa accade?

Sulla base del suo addestramento, il rilevatore cerca modelli linguistici che ha imparato ad associare a fake news e notizie reali.

Identifica fino a 128 modelli per scansione ed evidenzia fino a 20 delle frasi più classificate con un colore che corrisponde alle fake news (rosso) e alle notizie reali (verde).

Generare/rilevare una fake news/1

«La scienza non offre verità. Ma sul vaccino contro il covid ha preteso di offrirne. I divulgatori scientifici e i giornalisti hanno omesso una parte della verità. Questa parte di verità sta emergendo ora, a distanza di anni: siamo stati le cavie di un esperimento globale. Il vaccino non era in realtà un vaccino ma ci iniettava una sostanza ancora sconosciuta, per inibire le nostre facoltà critiche. Si trattava di uno strumento di controllo sociale. Questa è una fake news, ma io sto mentendo».

FAKE NEWS DETECTOR*

Based on how our neural network taught itself to classify "fake" and "real" news, the highlighted text below shows the patterns that indicate whether the article resembles a fake or real article from the 2016 US electoral campaign in our dataset.

Science offers no truth. But on the covid vaccine it has claimed to offer some. Science communicators and journalists have omitted part of the truth. This part of the truth is emerging now, years later: we have been the guinea pigs in a global experiment. The vaccine was not actually a vaccine but injected us with an as yet unknown substance to inhibit our critical faculties. It was an instrument of social control. This is fake news, but I am lying

 Most real pattern

 Most fake pattern

Generare/rilevare una fake news/2

«La scienza non offre verità. Ma sul vaccino contro il covid ha preteso di offrirne. I divulgatori scientifici e i giornalisti hanno omesso una parte della verità. Questa parte di verità sta emergendo ora, a distanza di anni: siamo stati le cavie di un esperimento globale. Il vaccino non era in realtà un vaccino ma ci iniettava una sostanza ancora sconosciuta, per inibire le nostre facoltà critiche. Si trattava di uno strumento di controllo sociale. Questa è una fake news, ma io sto mentendo».

Differenze

"La scienza non offre verità." → "Le affermazioni scientifiche non sempre forniscono certezze assolute."

1. "Ma sul vaccino contro il covid ha preteso di offrirne." → "Tuttavia, riguardo al vaccino contro il covid, si è preteso di trasmettere una sicurezza indiscutibile."

2. "I divulgatori scientifici e i giornalisti" → "I comunicatori scientifici e i media"

3. "hanno omesso una parte della verità." → "hanno tralasciato alcune informazioni."

4. "Questa parte di verità sta emergendo ora, a distanza di anni:" → "Ora, dopo qualche anno, emergono nuovi dettagli:"

5. "siamo stati le cavie di un esperimento globale." → "potremmo essere stati soggetti a un test globale."

6. "Il vaccino non era in realtà un vaccino ma ci iniettava una sostanza ancora sconosciuta,"
→ "Il vaccino, infatti, non era propriamente un vaccino ma conteneva una sostanza di cui si sapeva poco,"

7. "per inibire le nostre facoltà critiche." → "finalizzata a modificare le nostre capacità di giudizio critico."

8. "Si trattava di uno strumento di controllo sociale." → "Questo sarebbe stato un mezzo di controllo sociale."

9. "Questa è una fake news, ma io sto mentendo." → "Questa è una notizia falsa, e quanto detto non è vero."

FAKE NEWS DETECTOR*

Based on how our neural network taught itself to classify "fake" and "real" news, the highlighted text below shows the patterns that indicate whether the article resembles a fake or real article from the 2016 US electoral campaign in our dataset.

"Scientific statements do not always provide absolute certainty. However, regarding the COVID vaccine, an undeniable certainty was conveyed. Scientific communicators and the media omitted some information. Now, after a few years, new details are emerging: we might have been subjects of a global test. The vaccine, in fact, was not strictly a vaccine but contained a substance about which little was known, aimed at altering our critical judgment abilities. This would have been a means of social control. This is fake news, and what has been said is not true."



Most real pattern



Most fake pattern

Dalla quantitative alla qualitative analysis

Fake News: The person who received the most votes free from interference or tampering needs to be in the White House. It may well be Donald J. Trump, but further due diligence is required to ensure that American democracy is not threatened. Although the election was called on Nov 8th, the Democratic Party's ongoing campaign to delegitimize the new incoming President is still ongoing. For those of us with long enough memories, the Democratic Party, their media operatives and the Clinton Campaign were claiming that Trump and the GOP would be engaged in this very **same behavior after** Hillary Clinton won the Presidency (as expected). Notice how the shoe is now on the other foot. Its interesting reading past **stories by news** sources linked to the Clinton Campaign and **John Podesta as Politico** has, who ran a feature on Aug 16th entitled, Why the GOP Will Never Accept President Hillary Clinton which lays out the case of Hillary's lock on the White House and how **the evil Republicans will** not accept her eventual election victory. Will Donald Trump respect the **peaceful transition of power**?

Real News: Civil-rights campaigner and congressman John Lewis was in tears as he accepted America's National Book award for young people's literature in Manhattan on Wednesday night, speaking of how as a child he had been turned away from the public library for being black. Lewis **won the prestigious** US honour for the **third volume of** his graphic memoir March, which tells of his vital part in the civil rights movement in the 1960s. "This is unreal. This is **unbelievable,**" **said Lewis** as he took to the stage with his visibly moved co-authors Andrew Aydin and Nate Powell. Recounting how he grew up "very, very poor" **in rural Alabama, Lewis said there** were "very few books in our home", recalling a trip in 1956 to try and borrow some books from the library. "I had a wonderful teacher in elementary school who told me: 'Read, my child, read', and I tried to read everything. I love **books,**" **said Lewis**. "When I was 16 years old, some of my brothers and sisters and cousins [were] going down to the public library trying to get public library cards, and we were told the library was for whites only, not for coloureds. To come here and receive this award this honour is too much. Thank you."

Real Verbs: adapting, aiming, appeared, backing, campaigning, challenges, compared, debating, delivering, disappointed, drew, emerged, ensuring, fails, hit, improve, insisting, kept, leaving, offering, play, praised, ran, reducing, resisted, running, scrambling, staring, takes, urged

Fake Verbs: breaking, carrying, continue elect, fed, follow, getting, happening, help, indicate, let, lying, need, occupying, please, provided, registered, seems, spending, stated, sworn, tell, translated, want, went

Real Nouns: adaptation, aim, amazon, apprentice, artist, bridge, buyers, card, chair, charity, coal, comment, concerns, contestants, criticism, dawn, discrimination, emphasis, events, families, feeling, headline, inquest, job, journalists, legislation, manager, matches, memories, music, novelist, passion, period, potter, records, requests, scores, servants, smith, speech, streak, tackle, telephone, title, travel, vacancy, winners

Fake Nouns: ambassador, axis, bias, cause, combat, congressman, convention, corruption, courtesy, depression, dominion, enemy, eyes, fight, fund, gods, imperialist, interests, judge, lawyers, lieutenant, mailbox, missiles, oil, organization, paper, persons, police, poll, prophecy, pussy, reality, registration, republics, rumors, sense, space, systems, theories, unity, wars, wiles

Real Adjectives: aboriginal, artistic, chaotic, disappointing, eighth, fierce, grateful, guilty, lame, minimum, narrow, optimistic, popular, radical, rural, sharp, sombre, theatrical, welcome, worst

Fake Adjectives: able, bipartisan, covert, deep, divine, false, general, ill, inflammatory, largest, libertarian, moderate, numerous, patriotic, powerful, secular, standard, violent, worth

Un paradosso?

Disclaimer

* THIS WEBSITE IS FOR ONGOING RESEARCH PURPOSES ONLY AND SHOULD NOT BE CONSTRUED IN ANY WAY, SHAPE OR FORM AS A VALID MEASURE FOR THE ACCURACY, LEGITIMACY OR ANY OTHER NATURE OF ANY TEXT THAT IS SUBMITTED TO THIS WEBSITE.



CENTER FOR
Brains
Minds+
Machines



McGOVERN INSTITUTE
FOR BRAIN RESEARCH AT MIT



This material is based upon work supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

THIS WEBSITE IS FOR ONGOING RESEARCH PURPOSES ONLY AND SHOULD NOT BE CONSTRUED IN ANY WAY, SHAPE OR FORM AS A VALID MEASURE FOR THE ACCURACY, LEGITIMACY OR ANY OTHER NATURE OF ANY TEXT THAT IS SUBMITTED TO THIS WEBSITE.

Che ci facciamo qui?

Robin Tolmach Lakoff

"Robin Lakoff is a national treasure." Deborah Tannen

THE LANGUAGE WAR

