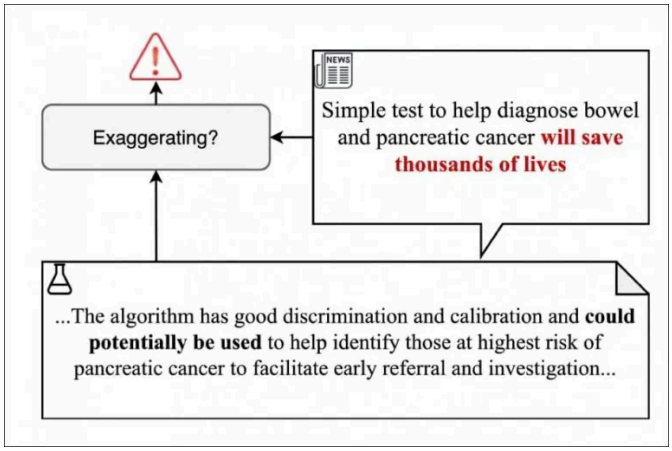
An NLP Approach to Exaggeration Detection in Science Journalism

Originally published at https://www.unite.ai/an-nlp-approach-to-exaggeration-detection-in-science-journalism/

Researchers from Denmark have developed an ‘exaggeration detection’ system designed to mitigate the effects of journalists over-stating the implications of new scientific research papers when summarizing and reporting them. The work has been prompted by the extent to which new published research into COVID-19 has been distorted in reporting channels, though the authors concede that it is applicable across a broad tranche of the general science reporting sector.

The paper, entitled *Semi-Supervised Exaggeration Detection of Health Science Press Releases*, comes from the University of Copenhagen, and notes that the problem is exacerbated by the tendency of publications to not include source links to the original research – an increasingly common journalistic practice that attempts to supplant the original paper and substitute the re-reported summary as ‘source knowledge’ – even where the paper is publicly available.



From the paper, a typical manifestation of exaggeration of scientific papers. Source: https://arxiv.org/pdf/2108.13493.pdf

The problem is not limited to external journalistic reaction to new papers, but can extend into other kinds of summary, including internal PR efforts of universities and research institutions; promotional material aimed at soliciting the attention of news outlets; and the useful referral links (and potential ammunition for funding rounds) that entail when journalists 'bite'.

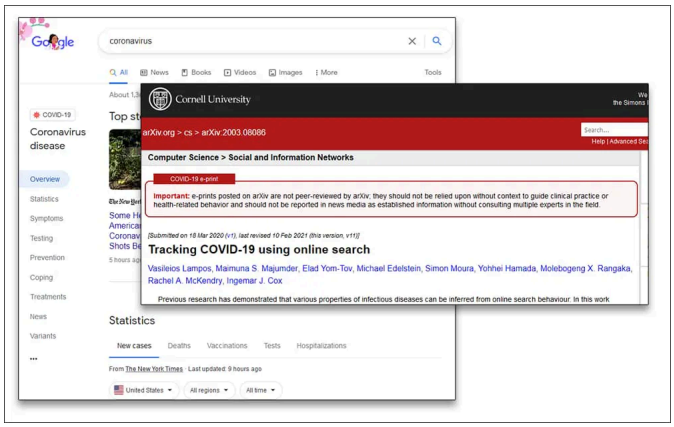
The work leverages Natural Language Processing (NLP) against a novel dataset of paired press releases and abstracts, with the researchers claiming to have developed '[a] new, more realistic task formulation' for the detection of scientific exaggeration. The authors have promised to publish the code and data for the work at GitHub soon.

Tackling Sensationalism

A number of studies have addressed the problem of scientific sensationalism over the last thirty or so years, and drawn attention to the misinformation that this can lead to. The late American scientific sociologist Dorothy Nelkin addressed the issue notably in the 1987 book *Selling Science: How the Press Covers Science and Technology*; the 2006 Embo report *Bad science in the headlines* highlighted the need for more scientifically-trained journalists, just as the internet was bringing critical budgetary pressures on the traditional media.

Additionally, in 2014 the British Medical Journal brought the problem into focus in a report; and a 2019 study from Wellcome Open Research even established that exaggeration of scientific papers confers no benefit (in terms of reach or traffic) to the news outlets and other reporting systems that perpetrate this practice.

However, the advent of the pandemic has brought the negative effects of this hyperbole into critical focus, with a range of information platforms, including the Google Search engine results page and Cornell University's Arxiv index of scientific papers now automatically adding disclaimers to any content that seems to be dealing with COVID.



Amended interfaces for searches and content related to COVID, from the Google search results page, and from Cornell University's influential Arxiv scientific paper repository.

Prior projects have attempted to create exaggeration detection systems for scientific papers by leveraging NLP, including a 2019 collaboration between researchers from Hong Kong and China, and another(unrelated) Denmark paper in 2017.

The researchers of the new paper note that these earlier efforts developed datasets of claims from abstracts and summaries from PubMed and EurekAlert, labeled for 'strength', and used them to train machine learning models to predict *claim strength* in unseen data.

MT-PET

The new research instead combines a press release and abstract as a combined data entity, and exploits the resulting dataset in MT-PET, a multi-task-capable version of the Pattern Exploiting Training research first presented in 2020 as *Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference*, a combined research effort from two German research institutions.

No existing dataset was found to be suitable for the task, and the team therefore curated a novel dataset of paired sentences from abstracts and related press releases, assessed by 'experts' in terms of their tendency to exaggerate.

The researchers used the few-shot text classification framework PETAL as part of a pipeline to automatically generate pattern-verbalizer pairs, subsequently re-iterating through the data until roughly equivalent tuplets were found for two qualities: exaggeration detection and claim strength.

The 'gold' data for testing was reused from the aforementioned earlier research projects, consisting of 823 pairs of abstracts and press releases. The researchers rejected the possible use of the 2014 BMJ data, since it is paraphrased.

This process obtained a dataset of 663 abstract/release pairs labeled for exaggeration and claim strength. The researchers randomly sampled 100 of them as few-shot learning training data, with 553 examples set aside for testing. Additionally a small training set was created consisting of 1,138 sentences, classified as to whether or not they represent the main conclusion of the summary or press release. These were used to identify 'conclusion sentences' in unlabeled pairs.

Testing

The researchers tested the approach in three configurations: a fully supervised setting with exclusively labeled data; a single-task PET scenario; and on the new MT-PET, which adds a secondary formulation thread as an auxiliary task (since the aim of the project is to examine two separate qualities from a dataset with paired data constructs).

The researchers found that MT-PET improved on the base PET results across testing environments, and found that identifying the claim strength helped to produce soft-labeled training data for exaggeration detection. However, the paper notes that in certain configurations among a complex array of tests, particularly related to claim strength, the presence of professionally labeled data may be a factor in improved results (compared to earlier research projects that address this problem). This could have implications for the extent to which the pipeline can be automated, depending on the data emphasis of the task.

Nonetheless, the researchers conclude that MT-PET *'helps in the more difficult cases of identifying and differentiating direct causal claims from weaker claims, and that the most performant approach involves classifying and comparing the individual claim strength of statements from the source and target documents'*.

In closing, the work speculates that MT-PET could not only be applied to a broader range of scientific papers (outside of the health sector), but could also form the basis of new tools to help journalists produce better overviews of scientific papers (though this, perhaps naively, assumes that journalists are exaggerating claim strength through ignorance), as well as aiding the research community in formulating a clearer use of language to explain complex ideas. Further, the paper observes:

> *'[it] should be noted that the predictive performance results reported in this paper are for press releases written by science journalists – one could expect worse results for press releases which more strongly simplify scientific articles.'*