



ORIGINAL ARTICLE

Automatically quantifying the scientific quality and sensationalism of news records mentioning pandemics: validating a maximum entropy machine-learning model

Steven J. Hoffman^{a,*}, Victoria Justicz^b

^aGlobal Strategy Lab, Faculty of Law, University of Ottawa, 57 Louis Pasteur St, Ottawa, Ontario, Canada K1N 6N5

^bHarvard Global Health Institute, Harvard University, 42 Church St, Cambridge, Massachusetts 02138, USA

Accepted 8 December 2015; Published online xxxx

Abstract

Objective: To develop and validate a method for automatically quantifying the scientific quality and sensationalism of individual news records.

Study design: After retrieving 163,433 news records mentioning the Severe Acute Respiratory Syndrome (SARS) and H1N1 pandemics, a maximum entropy model for inductive machine learning was used to identify relationships among 500 randomly sampled news records that correlated with systematic human assessments of their scientific quality and sensationalism. These relationships were then computationally applied to automatically classify 10,000 additional randomly sampled news records. The model was validated by randomly sampling 200 records and comparing human assessments of them to the computer assessments.

Results: The computer model correctly assessed the relevance of 86% of news records, the quality of 65% of records, and the sensationalism of 73% of records, as compared to human assessments. Overall, the scientific quality of SARS and H1N1 news media coverage had potentially important shortcomings, but coverage was not too sensationalizing. Coverage slightly improved between the two pandemics.

Conclusion: Automated methods can evaluate news records faster, cheaper, and possibly better than humans. The specific procedure implemented in this study can at the very least identify subsets of news records that are far more likely to have particular scientific and discursive qualities. © 2016 Elsevier Inc. All rights reserved.

Keywords: Health communication; Mass media; News; Pandemics; Machine learning; Validation studies

1. Introduction

The news media is one of the most powerful societal influences and most important sources of publicly available health information. It can significantly influence people's health-related behaviors [1], clinical practices [2], and policymaking processes [3]. Yet, current news coverage of health issues is not optimal. Prior studies have identified instances of health information being distorted or misreported in the news, presumably resulting in gaps among what researchers know about health issues, how journalists convey this information, and, ultimately, the reports on which

health professionals, policymakers, and the public act [4–7].

Accurate health news coverage is particularly important in the context of crises like pandemics, when events are rapidly unfolding, when facts are constantly changing, and when more credible sources may be unavailable or inaccessible [8]. But news coverage is probably no better during crises and may actually be worse. A recent systematic review, which integrated findings from 13 content-analytic studies, concluded that the news media may have dramatized the A/H1N1 influenza (H1N1) pandemic of 2009–2010 through excessive coverage and overemphasis on the threat posed by the virus rather than available self-protection measures [9]. Initial genomic studies of the H1N1 virus were reported sensationally and in isolation without being put in the context of the larger body of research to which they contributed. Worst-case scenarios for the H1N1 pandemic were sometimes laid out theatrically without caveating possible risks with any sense of

Conflicts of interest: We declare there are no competing interests.

Funding: S.J.H. is financially supported by the Canadian Institutes of Health Research, Research Council of Norway, and Trudeau Foundation.

* Corresponding author. Tel.: +1-613-562-5800x8811; fax: +1-613-562-5124.

E-mail address: steven.hoffman@uottawa.ca (S.J. Hoffman).

What is new?

Key findings

- A new automated method for quantitatively evaluating the relevance, scientific quality, and sensationalism of individual news records was developed and successfully modeled, applied, and validated on a huge corpus of news records mentioning two pandemics.

What this adds to what was known?

- Even rudimentary machine-learning models can accurately classify text documents for complex attributes like scientific quality and sensationalism.

What is the implication and what should change now?

- Automated text analysis and machine-learning modeling represent exciting frontiers in health research and news media analysis.
- With further developments, these approaches should be able to help detect performance gaps, identify problems, develop solutions, evaluate interventions, and hold organizations accountable.

the likelihood (or unlikelihood) in which they may or may not be realized. When high quality, specific information was available, the journalistic imperative of balanced coverage too often resulted in trustworthy evidence from credible scientists reported alongside ill-informed opinions from the most popular celebrities and conflicted lobbyists

[10]. Similar concerns were raised following the severe acute respiratory syndrome (SARS) outbreak in 2003 [11,12]. Likewise, the 2014 Ebola outbreak was consistently front-page news around the world for weeks—drawing unprecedented public interest (see Fig. 1)—despite only a single Ebola death outside of West Africa [13].

To researchers, this “research-to-reporting gap” and the broader “research-to-action gap” that it perpetuates is frustrating. But to those people who rely on the media as a primary source of health information—the health professionals who provide treatment, the policymakers who direct government action, and the public who make personal health decisions every day—this gap is potentially harmful. It means people may be routinely left to act on suboptimal information and unnecessary fear, and therefore cannot make informed decisions about how to respond to pandemics.

At the very least, suboptimal media coverage of pandemics reduces capacity to quickly access, assess, adapt, and apply emerging information as it is generated, disseminate public health guidance, and coordinate responses of health system stakeholders. More broadly, suboptimal coverage can diminish public discourse on policy issues, trust in science, and accountability for decisions, thereby affecting good governance, oversight and broader principles of civic engagement and democratic responsibility [3,14].

This study developed a systematic and comprehensive method for automatically quantifying the scientific quality and sensationalism of news media coverage which was then validated on a corpus of news records published during the SARS and H1N1 pandemic alert periods. Scientific quality is about accurate reporting that reflects truth and avoids bias [15]. Sensationalism is a discourse strategy of presenting news as more extraordinary, interesting, or relevant than is objectively warranted [16]. Analysis of vast quantities of qualitative data is aided by advances in automatic and

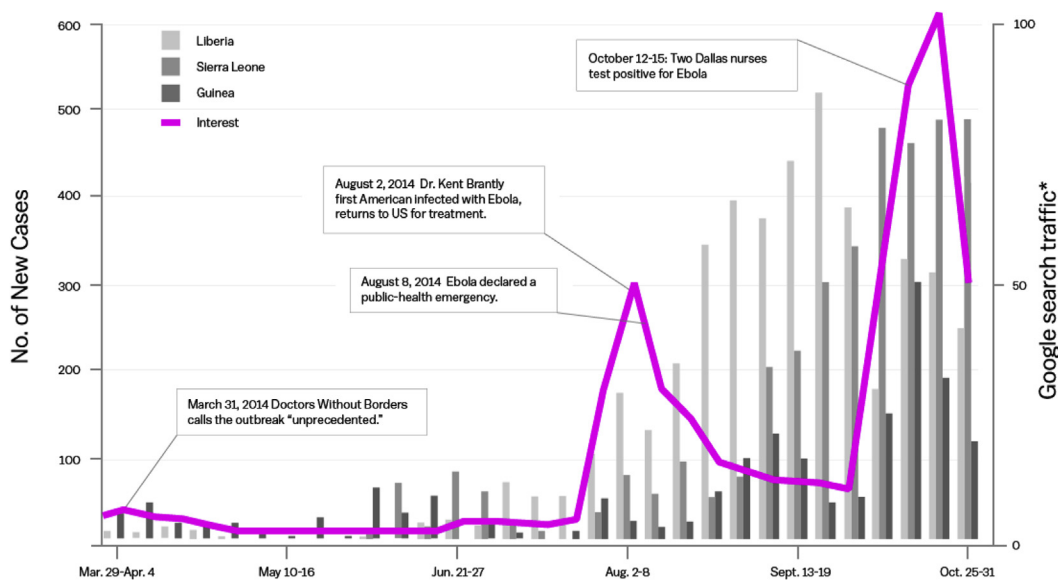


Fig. 1. Google searches for “Ebola” versus New Ebola Cases. Figure reproduced with permission from Vox Media Inc.

Download English Version:

<https://daneshyari.com/en/article/7519614>

Download Persian Version:

<https://daneshyari.com/article/7519614>

[Daneshyari.com](https://daneshyari.com)