Exploring informativeness of Clickbait headlines: You won't believe what we have found

MD MAIN UDDIN RONY, College of Information Studies, university of Maryland

1 PROPOSAL PROSE

Clickbait, a linguistic technique to grab readers' attention, has become a powerful tool in the highly competitive media market. Industry based on this practice is continuously and rapidly growing although it has some negative impacts on the media ecosystem. Clickbait draws some serious attention in the research community recently and many researchers put shed on this malpractice from different perspectives [4, 13, 15, 17]. Some researchers considered clickbait as an attention-grabbing and revenue-generating tool [14, 15, 17], some framed it as a viral marketing strategy [14], etc. Moreover, many works framed the problem from the psychological point of view of the readers [17], many studied it from social media perspectives [13] and some also analyzed it as a mode of communication [12]. As we can see Clickbait covers many disciplines in the existing literature, we need an interdisciplinary approach to understand it fully. Furthermore, an interdisciplinary characterization embracing most of the aspects (e.g. social, psychological, etc) may help address the unique challenges of automated detection that emerge from its diversity in terms of topics, styles, media platforms, etc. To this end, this paper explores one important characteristic of clickbait headline, Informativeness.

Although many researchers agreed that clickbait deliberately withholds important information to attract the readers clicking into a link [13, 14], the particular relationship between the informativeness and the level of clickbait is still unexplored. Although many researchers agreed that clickbait deliberately withholds important information to attract the readers clicking into a link [13, 14], the particular relationship between the informativeness and the level of clickbait is still unexplored. This study aims to exploit this gap of study based on a hypothesis that *the informativeness of a headline can explain its clickbait nature.*

1.1 Informativeness

Informativeness of a news headline can be defined as the amount of information it holds of its main content. Previously, many works explored the informativeness of a headline but nobody took the lead to define it formally. Growney et al. introduced the idea of informativeness while exploring the conditions that would promote older adults' attention to negative, but personally relevant information. Based on the presence of helpful information followed by a statement, the authors defined the level of informativeness (hi- and low-informative) [6]. Lopez et al. proposed a guideline to generate attractive headlines without the loss of informativeness and to get the cue of informativeness they relied on the noun phrases extracted from the headline [11]. In many headline generation tasks, informativeness has been used as a metric of quality [2, 10]. In the field

Author's address: Md Main Uddin Rony, College of Information Studies, university of Maryland, mrony@umd.edu.

XXXX-XXXX/2020/5-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

of journalism, 5W1H is a well-known concept to measure the informativeness. Hamborg et al. and Wei et al. used this concept to extract main events from a news headline [7, 18]. Wei et al. defined informativeness of the news headline based on this 5W1H concept. According to them, if the headline can answer the "who, what or whom" about a news event, it can be considered as informative [18]. In the current literature, there exists another concept to define informativeness, factual density. According to Sharevski [16], the number of facts contained in a document, normalized by its length can be an indicator of informativeness.

Horn et al. [8] and Lex et al. [9] also maintained the same concept of using factual density as a measurement of informativeness. Lex et al. curated a dataset of Wikipedia articles to check the quality of the web content using the factual density. Their experiment showed the efficacy of the factual density to measure the quality of the content. To capture the number of facts presented in the document, they used an open information extraction system. But open information extraction system may suffer from performance issues. Horn et al. extended the work of Lex et al. by incorporating human annotation to check the effectiveness of factual density. Moreover, they used arbitrary web content instead of Wikipedia articles only to cover the real internet environment. But both works used longer articles for their experiments. So, the efficacy of factual density to capture the informativeness of shorter sentences like headlines is still unexplored.

After exploring the current ground, factual density, and 5W1H are found to be more popular techniques to capture the informativeness. That's why, in this study, I am planning to use both of these approaches to representing the informativeness of the news headlines.

1.2 Level of Clickbaity-ness

Many studies considered clickbait as a binary decision (either clickbait or not) [1, 3, 15], but clickbait detection isn't a straightforward one and it can be considered as a subjective notion [13]. Chen et al. [5] claimed that people have an intuitive understanding of clickbait decision but it is mostly true for recognizing definite clickbait or non-clickbait classes. But there is a noticeable disagreement while deciding the middle level of clickbaity-ness. This decision behavior is also confirmed by Potthast et al. when they compiled a dataset, Webis Clickbait Corpus 2017. In this work, human annotators rated news headlines on a 4-point Likert scale that represents the four levels of clickbaity-ness (Not clickbaiting, Slightly clickbaiting, Considerably clickbaiting, Heavily clickbaiting). They found the interrater agreement fair and satisfying which allows us to consider the dataset as a reasonable representation of clickbait level.

1.3 Research Questions

Based on the conversation above, this study intends to establish a relationship between the clickbaity-ness and the informativeness of a headline solving two research questions.

- Is factual density an effective measurement to capture informativeness of news headlines?
- Can informativeness of a headline detect the level of clickbaity-ness of a news headline?
- How do users perceive the informativeness when they encounter different levels of clickbaity-ness?

REFERENCES

[1] Amol Agrawal. 2016. Clickbait detection using deep learning. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*. IEEE, 268–272.

[2] Enrique Alfonseca, Daniele Pighin, and Guillermo Garrido. 2013. Heady: News headline abstraction through event pattern clustering. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1243–1253.

[3] Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer. 2016. " 8 Amazing Secrets for Getting More Clicks": Detecting Clickbaits in News Streams Using Article Informality. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[4] Chris Brogly and Victoria L Rubin. 2018. Detecting Clickbait: Here's How to Do It/Comment détecter les pièges à clic. *Canadian Journal of Information and Library Science* 42, 3 (2018), 154–175.
[5] Yimin Chen and Victoria L Rubin. 2017. Perceptions of Clickbait: A Q-Methodology Approach. In *Proceedings of the 45th Annual Conference of The Canadian Association for Information Science/L'Association canadienne des sciences de l'information (CAIS/ACSI2017), Ryerson University, Toronto, May 31-June 2, 2017*.
[6] Claire M Growney and Thomas M Hess. 2019. The influence of mood versus relevant self-perceptions in older adults' interest in negative health-related information. *Psychology and aging* (2019).
[7] Felix Hamborg, Corinna Breitinger, Moritz Schubotz, Soeren Lachnit, and Bela Gipp. 2018. Extraction of main event descriptors from news articles by answering the journalistic five W and one H questions. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. 339–340.
[8] Christopher Horn, Alisa Zhila, Alexander Gelbukh, Roman Kern, and Elisabeth Lex. 2013. Using factual density to measure informativeness of web documents. In *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)*. 227–238.
[9] Elisabeth Lex, Michael Voelske, Marcelo Errecalde, Edgardo Ferretti, Leticia Cagnina, Christopher Horn, Benno Stein, and Michael Granitzer. 2012. Measuring the quality of web content using factual information. In *Proceedings of the 2nd joint WICOW/AIRWeb workshop on web quality*. 7–10.
[10] Marina Litvak, Natalia Vanetik, and Itzhak Eretz Kdosha. 2019. HEvAS: Headline Evaluation and Analysis System. In *Proceedings of the Workshop MultiLing 2019: Summarization Across Languages, Genres and Sources*. 73–80.
[11] Cédric Lopez, Violaine Prince, and Mathieu Roche. 2014. How can catchy titles be generated without loss of informativeness? *Expert systems with applications* 41, 4 (2014), 1051–1062.
[12] Dolors Palau-Sampio et al. 2016. Reference press metamorphosis in the digital context: clickbait and tabloid strategies in Elpais. com. *Communication & Society* 29, 2 (2016), 63–79.
[13] Martin Potthast, Tim Gollub, Matthias Hagen, and Benno Stein. 2018. The clickbait challenge 2017: towards a regression model for clickbait strength. *arXiv preprint arXiv:1812.10847* (2018).
[14] Martin Potthast, Tim Gollub, Kristof Komlossy, Sebastian Schuster, Matti Wiegmann, Erika Patricia Garces Fernandez, Matthias Hagen, and Benno Stein. 2018. Crowdsourcing a large corpus of clickbait on twitter. In *Proceedings of the 27th International Conference on Computational Linguistics*. 1498–1507.
[15] Md Main Uddin Rony, Naeemul Hassan, and Mohammad Yousuf. 2017. Diving deep into clickbaits: Who use them to what extents in which topics with what effects?. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. 232–239.
[16] Filipo Sharevski. [n.d.]. THE APPROPRIATENESS OF THE FACTUAL DENSITY AS AN INFORMATIVENESS MEASURE FOR ONLINE NEWS. ([n. d.]).
[17] Neoh Sun Yi Vincent, Anjan Pal, and Alton YK Chua. 2018. Studying Healthcare Personnel's Intention to Click Clickbaits. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. 1.
[18] Wei Wang. 2012. Chinese news event 5W1H semantic elements extraction for event ontology population. In *Proceedings of the 21st International Conference on World Wide Web*. 197–202.