Clickbait Spoiling using Question Answering models

Ankur Chemburkar **Aravind Krishnan** **Nehal Muthukumar**

Siraj Sandhu **Somil Kiran Jain**

University of Southern California
Viterbi School of Engineering

1 Project Domain and Goals

We are addressing the problem of clickbait spoiling - generating a short text that satisfies the curiosity generated by a clickbait post. The task is to understand the question posed by the clickbait, summarize the text on the linked webpage by locating essential text spans and extract or generate a concise, informative response that eliminates the need to click on the link. Any model tackling this task requires various elements of Natural Language Processing like text understanding and text generation. This problem exists in the intersection of question-answering and text summarization subdomains of NLP. We use question-answering to address this task by considering the clickbait as the 'query' and the linked article as the 'context' to extract the answer.

Clickbait headlines often use sensational language, misleading information, or exaggerated claims to provoke curiosity and compel readers to click on the link. The conclusion of the linked webpage is also usually less exciting than the clickbait portrays. Clickbaits have become a pervasive issue on the internet that raises ethical concerns, as they can manipulate readers' emotions and exploit their vulnerabilities for financial gain. In addition, clickbait can contribute to spreading fake news and disinformation, as it can lure readers into clicking on links that offer misleading or biased content. To address this issue, we generate spoilers that can quench the curiosity induced by clickbait posts.

This project aims to provide users with an efficient way to avoid clickbait articles that often disappoint or waste time by providing misleading or trivial information. The project also has the potential to promote ethical and honest content creation by reducing the incentives for clickbait tactics.

2 Related Work

Clickbait detection technologies were originally proposed by Rubin et al. (2015), and the first two detectors were independently developed by Potthast et al. (2016) (random forest classifier) and Chakraborty et al. (2016) (SVM with RBF kernel), both of which employ extensive feature engineering spanning the structural, word level, N-gram, and linguistic categories. Agrawal (2016) presented a convolutional neural network for the classification of headlines containing clickbait. Since then, Hagen et al. (2022) has observed that transformer models have dominated this task.

Hagen et al. (2022) introduced a corpus of 5000 annotated resolved/spoiled posts and concluded that the question-answering model DeBERTa-large (He et al., 2021) outperforms all others in generating spoilers for both types of spoilers needed, i.e., phrase or passage. Hättasch et al. (2020) also introduced a corpus comprising 2635 annotated samples for English language clickbait articles and analyzed the data's usefulness using several baseline approaches. Of these, the best performing one was the T5 (Raffel et al., 2022), finetuned first on the SQuAD (Rajpurkar et al., 2016) dataset and then on their augmented corpus. The augmentation used NLPAug (Ma, 2019) to create two artificial teasers for each training sample based on WordNet.

In considering the problem as a question-answering problem, two benchmark datasets considered by Hagen et al. (2022) were (1) SQuAD (Rajpurkar et al., 2016), which compiles 107,785 questions and answers based on 536 Wikipedia articles—93.6% of which are factual in nature, such as names, noun phrases and numbers, and (2) TriviaQA (Joshi et al., 2017), which contains 95,000 question–answer pairs based on challenging trivia. The models used in their experiments were chosen for their (then) state-of-the-art performance on the above benchmarks.

A limitation of the question-answering approach is the typically brief nature of "answers" in the datasets described above. Framing the problem then as passage retrieval, which allows longer passages (e.g., one or more sentences) of text as answers, can be thought to solve the issue. One benchmark is the dataset for a large question-answering task proposed by Bajaj et al. (2016), with 8.8 million passages derived for a set of 100,000 questions originally submitted to Bing. Guo et al. (2020) and Lin et al. (2020) observed in their survey that neural retrieval models have been successful in the passage retrieval problem. However, Hagen et al. (2022) concluded there was no significant uplift in the generation of passage spoilers compared to phrasal spoilers by using this approach, and overall is drastically worse. They obtained the best results with the MonoT5 model (Nogueira et al., 2020).

Another way to formulate the problem can be topic-focused single document summarization. Narayan et al. (2018) introduced an abstractive model conditioned on the document's topics and based on convolutional neural networks. Deutsch and Roth (2019) studied a two-step abstractive model that first extractively selects a small number of sentences and then abstractively summarizes them. They observed that topical and partial summaries in the steps help the models identify relevant content.

3 Datasets

Our project will use the existing dataset Webis-Clickbait-22, which is publicly available at `https://webis.de/data/webis-clickbait-22.html`.

The dataset is in JSON Lines format (.jsonl), with each line containing a clickbait post, manually cleaned versions of the linked documents, and extracted spoilers for each clickbait post. The dataset includes fields such as postText, targetParagraphs, targetTitle, humanSpoiler, spoiler, spoilerPositions, and tags, which will be used in our project for tasks such as spoiler type classification and spoiler generation.

Some of the essential fields in the dataset include:

postText The text of the clickbait post that needs to be spoiled

targetParagraphs The paragraphs of manually extracted content of the linked web page used to classify the spoiler type and to generate the spoiler

targetTitle The title of the linked web page

humanSpoiler The abstractive spoiler generated by a human for the clickbait post from the linked web page (only available in training and validation datasets)

tags The spoiler type (phrase, passage, or multi) used to classify the spoiler

Since the dataset has already been preprocessed, we will not need further cleaning or processing. However, we are considering additional preprocessing techniques to improve the model's performance. We may also use SQuAD (Rajpurkar et al., 2016) and TriviaQA (Joshi et al., 2017) for pre-training the models. Our primary focus will be experimenting with different architectures to classify and generate spoilers for clickbait posts effectively.

4 Technical Challenges

Our main challenge would be to design a generative model that successfully summarizes the article's content. The spoiler-type classification can be either a phrase or a passage, depending on the content, for which we will build a classifier model. We believe that solving this pseudo task can lead to substantial results. Based on the spoiler type, the appropriate model will generate the spoiler to negate the clickbait. Clickbait spoiling has two approaches, question answering and passage retrieval. We will evaluate which approach provides the best results. Question answering has been proven to be a better approach in related works. We will verify their conclusions and build our model based on that.

Clickbait Challenge at SemEval 2023 has developed a baseline for evaluation, which we will use as a benchmark for measuring the performance of our model. To evaluate the effectiveness of spoiler type classification on such clickbait posts, we conduct three experiments: (1) multi-class, (2) one-vs-rest, and (3) one vs-one for the types of phrase and passage spoilers as done in previous papers. The accuracy they achieved was 80% which we will try to improve. To assess the quantitative correspondence between a derived spoiler and the ground truth, we shall use three question-answering oriented and one passage retrieval-oriented measure: BLEU-4 (Papineni et al., 2002), METEOR (Denkowski and Lavie, 2014), BERTScore (Zhang et al., 2020), and Precision@1.

5 Division of Labour

Phase 1:
Ankur: Data preparation, baseline model, and evaluation. Aravind, Nehal and Siraj: Replicate the results from the paper by running three models each. Somil: Experiment on abstractive text summarization for clickbait.
Phase 2:
Team: Discuss and brainstorm ideas for document engineering and generating spoilers. Aravind: Implement and test the idea of identifying multipart spoilers via the document structure. Ankur and Nehal: Generate phrase, passage, and multipart spoilers using a wide range of approaches for task 2 and test them. Siraj and Somil: Use Query Performance Prediction to identify the most promising spoiler type and to show spoilers only when we are confident that they are correct. Team: Explore and apply other query understanding approaches, such as pre-trained models from query intent prediction and query reformulation, to improve the performance of the system.

References

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

Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. 2016. Ms marco: A human generated machine reading comprehension dataset.

Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 9–16.

Michael Denkowski and Alon Lavie. 2014. Meteor universal: Language specific translation evaluation for any target language. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 376–380, Baltimore, Maryland, USA. Association for Computational Linguistics.

Daniel Deutsch and Dan Roth. 2019. Summary cloze: A new task for content selection in topic-focused summarization. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3720–3729, Hong Kong, China. Association for Computational Linguistics.

Jiafeng Guo, Yixing Fan, Liang Pang, Liu Yang, Qingyao Ai, Hamed Zamani, Chen Wu, W. Bruce Croft, and Xueqi Cheng. 2020. A deep look into neural ranking models for information retrieval. *Information Processing Management*, 57(6):102067.

Matthias Hagen, Maik Fröbe, Artur Jurk, and Martin Potthast. 2022. Clickbait spoiling via question answering and passage retrieval. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7025–7036, Dublin, Ireland. Association for Computational Linguistics.

Benjamin Hättasch, Nadja Geisler, Christian M. Meyer, and Carsten Binnig. 2020. Summarization beyond news: The automatically acquired fandom corpora. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6700–6708, Marseille, France. European Language Resources Association.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. {DEBERTA}: {DECODING}-{enhanced} {bert} {with} {disentangled} {attention}. In *International Conference on Learning Representations*.

Mandar Joshi, Eunsol Choi, Daniel Weld, and Luke Zettlemoyer. 2017. TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1601–1611, Vancouver, Canada. Association for Computational Linguistics.

Jimmy Lin, Rodrigo Nogueira, and Andrew Yates. 2020. Pretrained transformers for text ranking: Bert and beyond.

Edward Ma. 2019. Nlp augmentation. https://github.com/makcedward/nlpaug.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics.

Rodrigo Nogueira, Zhiying Jiang, Ronak Pradeep, and Jimmy Lin. 2020. Document ranking with a pretrained sequence-to-sequence model. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 708–718, Online. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. 2016. Clickbait detection. In *Advances in Information Retrieval*, pages 810–817, Cham. Springer International Publishing.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2022. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1).

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas. Association for Computational Linguistics.

Victoria Rubin, Niall Conroy, and Yimin Chen. 2015. Towards news verification: Deception detection methods for news discourse. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS48) Symposium on Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium, Kauai, Hawaii, USA*.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with BERT. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.