Clickbait Detection in Twitter

Tianyu Yang
Georgetown University
Washington, U.S.
ty233@georgetown.edu

Benita Neo Yan Ting
Georgetown University
Washington, U.S.
bn203@georgetown.edu

Introduction

With the proliferation of social media and online news media as a platform for news and gossips, today we are surrounded by unreliable content that has gone unchecked and unregulated. The competition for reader attention is rife, since readership translates directly to site popularity and revenues. Therefore many online media sites make use of clickbait as a tool to stand out from others. Clickbait articles are articles with catchy headlines with the intention of drawing attention of the readers, but their content do not live up to the hype of the headlines, or may be irrelevant to the article headlines. Taboola, one of the key providers of clickbait content on the web claims that its monthly reach of 500 million unique users doubled to 1 billion within a year. Clickbait taints the quality and role of online media sites providing news to readers, and violates journalistic codes of ethics. It boils down to each readers discretion, scrutiny, and knowledge to discern if the content is true, factual and up-to-date.

Clickbait works because it exploits the cognitive phenomenon (Curiosity Gap), where headlines forward referencing cues aim to generate enough curiosity so readers are compelled to click on the links to fill the gap in knowledge. These article headlines are optimized in real-time - various headlines are tried and tested at the same time, and the ones with the highest number of click-throughs will be the successful headlines used.

Currently, social media sites like Facebook promised toimporve on clickbait detection algorithms, using click-to-share ratio data and the amount of time readers spend on these articles.

The key purpose of our project is to quantify the extent to of clickbait tweets and to detect clickbait articles on Twitter. Given a tweet, we want to find out how well do our classifiers fare in detecting if it is a clickbait tweet. Our dataset is taken from the Clickbait Challenge 2017 data, containing a total of 19538 tweets, and 4761 tweets are recorded to be clickbait (25%), while the other 14777 are recorded to be non-clickbait tweets.

Related Works

For our project we referenced 3 research papers that were very relevant to our topic of clickbait in online news media.

The first is Stop Clickbait Detection and Preventing Clickbait in Online News Media. The dataset used in this paper is extenseive, with 18513 non-clickbait headlines extracted from wikinews articles, on the basis that these articles are considered gold standard for non-clickbaits, because of the rigorous checks and verification before publication. This project uses Downworthy rule as the baseline, which means detecting clickbait headlines using a fixed set of common clickbait phrases. The classifier built uses linguistic features of headlines like length of sentences, syntactic dependencies, distribution of stopwords, part-of-speech etcetera. This project also provided user-level blocking, customized based on the users browsing history by topic similarity and linguistic patterns.

The second paper is called From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles, focusing on the mismatch of headlines and content by analyzing the TF-IDF scores determined by n-grams in titles and articles. It used logistic regression to produce a multi-class regression: agree, disagree, discuss, indicating the stance of the headline towards the content of the article.

The third paper is Clickbait Detection, which focuses on tweets, and therefore has twitter-specific characteristics like hashtags, mentions, links, meta-information like media attachment and timestamp. Its data is randomly sampled from the top 20 most prominent publishers on Twitter according to retweets.This paper also uses downworthy as a baseline for comparison. It is manually assessed that 26% of twitter articles are considered clickbait. In this paper, it is also found that Listicles have a high probability of being clickbaits. Listicles are content presented in the form of a list, like 21 photos that will restore your faith in humanity, or 5 reasons why wine is good for your mental health. Classifier models used in this paper are logistic regression, naive bayes and random forest.

Data and Methodology

We used the 2017 clickbait challenge dataset, which includes 170331 records of tweet that provide link to another article. Each record contains contains id, timestamp of the post, text of the tweet, image(if any) in the tweet, title, description, keywords, content and caption of images(if any) of the linked article. The labels of a record includes a binary class label(clickbait or not clickbait) and probability of being clickbait provided by five annotators.

Among all text fields, tweets deserve the most attention for preprocessing as people tend to write informally and use various abbreviations, emoticons and emojis. Other text fields such as article title, description, have a more formal structure and use of language for both clickbait and nonclickbait.

The procedures of classification follows the guidance from [1] and [2]. After preprocessing all relevant text fields, we extracted potentially useful features covering the linguistic characteristic and apply machine learning models on the features

Experimental Plan and Evaluation

Most of the extracted features are from tweets and article title as they are the more important aspect. Features cover length of the text, distribtution of stop words, usage of punctuation, usage of twitter specific language such as hashtag, mentions, emoticons and emojis. Sentiment and subjectivity are measured using textblob, a lexicon based approach, in python. Other linguistic features include usage of propernoun, cardinal number and superlative adjective. In [3], the authors addressed clickbaits from the perspective of fake news, often appears in the form of title and content mismatch. In domain of tweets, 'fake news' can appear as content mismatch between post content and targeted article. For example, one of the clickbait instance in our data has post content "this is good" and article title relating to general election. To address this issue, we measured the n-gram jaccard similarity between tweet and article title with n up to 5.

Preprocessing involving replacing emoticons and emojis with tags, normalizing abbreviations. To address the discrepancy of writing styles between tweet, article title, description and keyewords, we used four count vectorizers with different n-gram level.

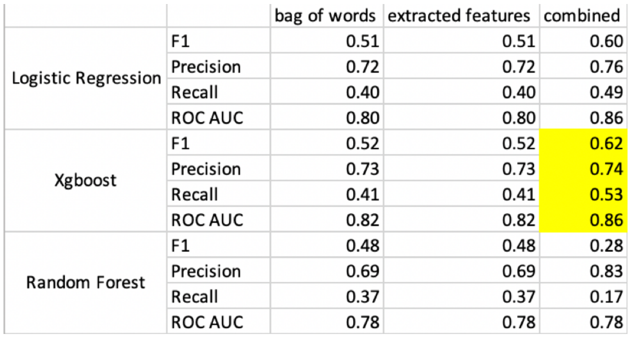
We started with binary classification based on the extracted features and bag of words feature, compared and analyzed results with precision, recall, f1 and roc-auc using logistic regression, xgboost and random forest. Extracted features were evaluated with stand alone performance when added to the models.

Since the data provide both label and annotated probabilities, to better reflect personal judgements, we created multiclass labels based on the average of annotated judgment. "Not Clickbait" covers mean probability from 0 to 0.33, "Maybe Clickbait" ranges from 0.33 to 0.66, and the rest is 'Definite Clickbait'. In addition to models used in binary classification, we experimented with recurrent convolutional network with glove word embedding. The construction of the model follows [4], where text input are transformed into vectors according to the embedding, fed into a layer of bidirectional LSTM cells, and connected with output through global maximum pooling. The evaluation metrics are precision, recall, roc-auc and f1 for individual classes.

All metrics recorded are average of three-fold cross validation results.

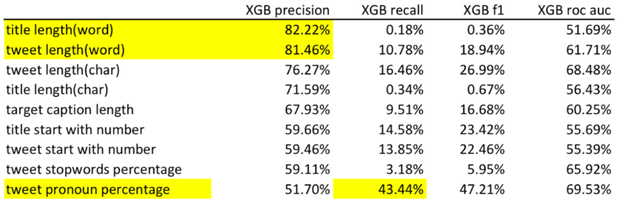
Results and Analysis

Figure 1 shows the best performing model,xgboost, for the binary classification task. The scores are in close range with the best performing model in the 2017 clickbait challenge. However, the classification performance is significantly worse than [1], where the authors focus on clickbaits on webpages instead of social platforms such as twitter. This suggest that overall, identifying clickbaits with more complex information structure is a difficult task.



		bag of words	extracted features	combined
Logistic Regression	F1	0.51	0.51	0.60
	Precision	0.72	0.72	0.76
	Recall	0.40	0.40	0.49
	ROC AUC	0.80	0.80	0.86
Xgboost	F1	0.52	0.52	0.62
	Precision	0.73	0.73	0.74
	Recall	0.41	0.41	0.53
	ROC AUC	0.82	0.82	0.86
Random Forest	F1	0.48	0.48	0.28
	Precision	0.69	0.69	0.83
	Recall	0.37	0.37	0.17
	ROC AUC	0.78	0.78	0.78

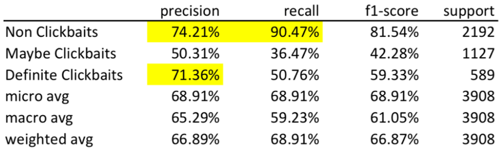
Fig. 1. binary classification result



	XGB precision	XGB recall	XGB f1	XGB roc auc
title length(word)	82.22%	0.18%	0.36%	51.69%
tweet length(word)	81.46%	10.78%	18.94%	61.71%
tweet length(char)	76.27%	16.46%	26.99%	68.48%
title length(char)	71.59%	0.34%	0.67%	56.43%
target caption length	67.93%	9.51%	16.68%	60.25%
title start with number	59.66%	14.58%	23.42%	55.69%
tweet start with number	59.46%	13.85%	22.46%	55.39%
tweet stopwords percentage	59.11%	3.18%	5.95%	65.92%
tweet pronoun percentage	51.70%	43.44%	47.21%	69.53%

Fig. 2. Feature Performance

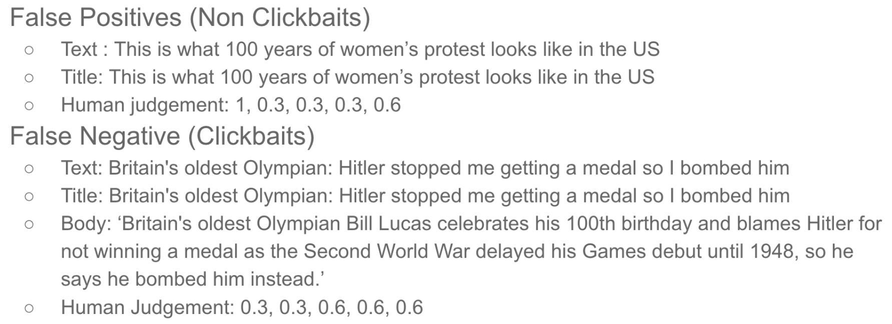
The error instances are made up with tweets that are confusing in nature. Figure 4 shows two typical example. The false positive example has sentence structure that largely matchs the characteristics of clickbaits. The sentence starts with a pronoun and does not contain many content words. Even the annotators do not have unifying opinions on the instance. On the contrary, the false negative case possesses many features of news articles and the annotators are not agreeing on the matter. Figure 2 shows that the important stand alone features are text length and pronoun distribution.



	precision	recall	f1-score	support
Non Clickbaits	74.21%	90.47%	81.54%	2192
Maybe Clickbaits	50.31%	36.47%	42.28%	1127
Definite Clickbaits	71.36%	50.76%	59.33%	589
micro avg	68.91%	68.91%	68.91%	3908
macro avg	65.29%	59.23%	61.05%	3908
weighted avg	66.89%	68.91%	66.87%	3908

Fig. 3. Multiclass classification performance

Figure 3 shows the classification performance of xgboost(best performing) for multiclass classification.The results suggest that 'Maybe Clickbaits' composed of confusing instances contributed to the overall difficulty of the clickbait detection. Figure 5 shows that error instances are so confusing that annotators starting to have bipolar opinions. However, in clickbait blockage applications, 'Maybe Clickbaits' performance does not create too many practical concerns as it can be



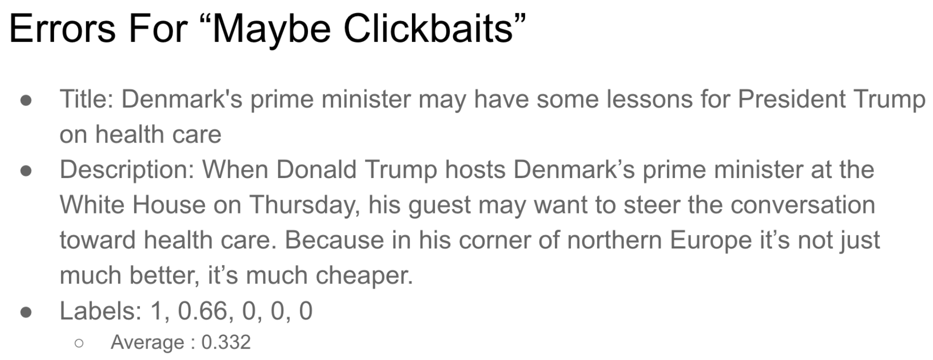
False Positives (Non Clickbaits)

- Text : This is what 100 years of women’s protest looks like in the US
- Title: This is what 100 years of women’s protest looks like in the US
- Human judgement: 1, 0.3, 0.3, 0.3, 0.6

False Negative (Clickbaits)

- Text: Britain's oldest Olympian: Hitler stopped me getting a medal so I bombed him
- Title: Britain's oldest Olympian: Hitler stopped me getting a medal so I bombed him
- Body: ‘Britain's oldest Olympian Bill Lucas celebrates his 100th birthday and blames Hitler for not winning a medal as the Second World War delayed his Games debut until 1948, so he says he bombed him instead.’
- Human Judgement: 0.3, 0.3, 0.6, 0.6, 0.6

Fig. 4. Binary classification Errors



Errors For “Maybe Clickbaits”

- Title: Denmark's prime minister may have some lessons for President Trump on health care
- Description: When Donald Trump hosts Denmark’s prime minister at the White House on Thursday, his guest may want to steer the conversation toward health care. Because in his corner of northern Europe it’s not just much better, it’s much cheaper.
- Labels: 1, 0.66, 0, 0, 0
  - Average : 0.332

Fig. 5. Multiclass classification error

combined with users perference. If an instance is classified as ’Maybe Clickbait’, user can choose to block. Further analysis and models and be build for this specific class since it normally has confusing structures.

Conclusion

This proeject provide an practical approach for identifying clickbaits on social platforms like twitter and provide insights on the discrepancy of performance of clickbait detection on general webpages and social platforms. Information in social platform possesses more complex structure and therefore allows clickbait creators to devise various forms of clickbait.

During the end of the project, we noticed that many errors in multiclass classificatino results are related to political topics. Recall that the hazzard of clickbait articles comes from offering less information than the titles guarantee. In some cases, although the title is as intriguing as clickbait title, the content provides enough information. This suggests that restricting merely at title, tweet structures and other linguistics characteristics provides insufficient aspects of anaysis. Future research should also include analysis on the content and background information of the target to generate an inclusive solution.

References

[1] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop clickbait: detecting and preventing clickbaits in online news media. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM ’16). IEEE Press, Piscataway, NJ, USA, 9-16.

[2] Potthast M., Kpsel S., Stein B., Hagen M. (2016) Clickbait Detection. In: Ferro N. et al. (eds) Advances in Information Retrieval. ECIR 2016. Lecture Notes in Computer Science, vol 9626. Springer, Cham

[3] Bourgonje, Peter & Moreno Schneider, Julian & Rehm, Georg. (2017). From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles. 84-89. 10.18653/v1/W17-4215.

[4] LAI, S.; XU, L.; LIU, K.; ZHAO, J.. Recurrent Convolutional Neural Networks for Text Classification. AAAI Conference on Artificial Intelligence, North America, feb. 2015