Controllable Clickbait Text Generation

Dusan Svilarkovic, Dominik Stammbach, Afra Amini, Elliott Ash
ETH Zürich

1 Introduction

In the previous 5 years, thanks to the development of attention-based transformers (Vaswani et al., 2017) and its encoder (Devlin et al., 2018), decoder (Radford et al., 2019) and encoder-decoder variations (Lewis et al., 2019; Zhang et al., 2019), NLP has become a rapidly developing field, with a lot of computational power spent on pre-training multi-billion parameter models, currently only being able to be trained with the resources held by the private companies, and only recently by the public organization created by the open initiatives (Big Science, 2021). Due to computational capability requirements, the majority of the research has been focused on "tweaking" the currently existing state-of-the-art models. Model "tweaking" has become a challenge to overcome in controlled, topic-based text generation settings in selected approaches: CTRL (Keskar et al., 2019) Fudge (Yang and Klein, 2021) and PPLM (Dathathri et al., 2019).

Although transformer architectures have been capable of generating text through conditioned attribute decoding with previously mentioned approaches (CTRL (Keskar et al., 2019), Fudge (Yang and Klein, 2021), PPLM (Dathathri et al., 2019)), it was shown to be working with changing performance on explicit attributes. So far, no one has attempted yet to understand how to gradually increase attribute grade while generating text. Moreover, there have been attempts to conditionally, based on meta-attributes generate text for the newspaper articles, such as in Grover (Zellers et al., 2019), and while it has given excellent results and has proven to be able to create humanly-reasonable text, it still had a structure constraint and had no control over how conditioned meta-attribute might appear in the text.

Inspired by the previous conditioned approaches to generating fake news articles (Zellers et al., 2019), and their worries that their and existing models might be capable of generating synthetic articles, we are proposing to create a controllable, attribute-based conditioned model for generating text based on the particular attribute. In our approach, we are generating clickbait titles, conditioned on the news article content, based on the "clickbait" grade of our classifier, and explore how this can be controlled through hyperparameter search. Moreover, unlike PPLM (Dathathri et al., 2019) and Fudge (Yang and Klein, 2021), which are using no-context, decoder-based GPT-2 (Radford et al., 2019), in our case, for generating news article titles, we use news article content and encoder-decoder based transformers (Zhang et al., 2019), (Beltagy et al., 2020) to generate titles.

2 Theoretical background

If we use probabilistic terminology, our goal is to create controllable, on-attribute conditioned text generation model $P(X|a, X')$, where X is a natural language sequence (news article title), a is an attribute, in our case "clickbaitiness", and X' is another natural language sequence (news article content). To create that model, we need a large language model $P(X|X')$ and an on-attribute trained classifier $P(X|a, X')$. By using the unnormalized Bayes rule, we create a conditional text-to-title generator

$$P(X|a, X') = P(a|X, X') \times P(X|X').$$

For using this model, only on-attribute classifier needs to be trained.

3 Data and Summary Statistics

For training on-attribute classifier, we use Webis dataset (Potthast et al., 2018), a clickbait dataset, containing a single score label from 0 to 1 by a step of 0.33 for the how "clickbaity" a title is. Important attributes per sample are title, article content, and clickbait label. Dataset has 38,517 samples separated into the train, validation, and test datasets.

The clickbait truth label is binary values (0 for no-clickbait, 1 for clickbait). Webis dataset was created with the help of 5 MTurk workers, for each article giving a grade on a 4 values scale from 0 to 1 by 0.33 step (not-clickbait (0.0), slightly clickbait (0.33), considerably clickbait (0.66), heavily clickbait (1.0)), and the binary truth label was given from the mean and rounding an average grade. In total, 9,276 posts are clickbait. In figure 1, we are showing a screenshot of a comparison of various previously existing datasets. Usage of previously created datasets was abandoned, due to the lack of meta-attributes in those datasets and the possibility of overlapping news articles and data leakage in train/val/test splits with Webis dataset.

Publication	Genre	Teaser	Acquisition	Annotation (Scale, People, κ)			Articles	Size
Agrawal (2016)	News	Headline	Gatekeeper-based	Binary	3	0.84	No	2,388
Potthast et al. (2016)	News	Tweet	Importance-based	Binary	3	0.35	Yes	2,992
Biyani et al. (2016)	News	Headline	Reputation-based?	Binary	?	?	?	4,073
Chakraborty et al. (2016)	News	Headline	Reputation-based	Binary	3	0.79	No	15,000
Rony et al. (2017)	News	Headline	Reputation-based	Binary	3	0.79?	No	32,000
Webis-Clickbait-17	News	Tweet	Importance-based	Graded	5	0.36	Yes	38,517

Table 1: Overview of clickbait corpora. All are in English; question marks (?) indicate lack of clarity.

Figure 1: A dataset table taken from Webis paper (Potthast et al., 2018). It shows how Webis-clickbait is the largest clickbait title dataset so far, and contains the largest scale range compared to the other datasets.

Furthermore, we also provide dataset content in figure 2, which shows the frequency of clickbait grades among the different publishers, showing that for example, BuzzFeed and Breitbart's news has the highest clickbait content out of all publishers, while New York Times and Guardian are more on the opposite side of the clickbaitiness spectrum.

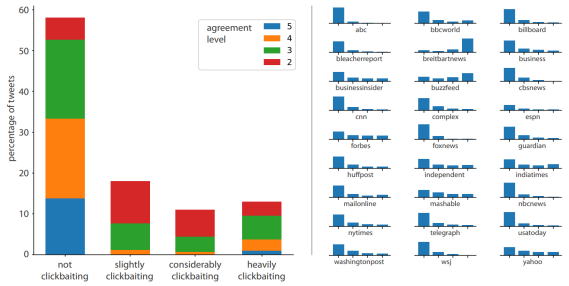

Figure 2: A dataset content per publisher and by clickbait grade.Taken from Webis paper (Potthast et al., 2018)

For further comparison of per grade scale title contents, we have also created word clouds for each grading scale and they are shown on a figure 3. As you can see, clickbaity titles contain some phrases like "you need to know" or "these are the # examples", meaning that there exist specific patterns in the title creation that can attract the readers, and that it is appearing in our dataset. For table-based visualization of 3-grams, please check out the Appendix A.

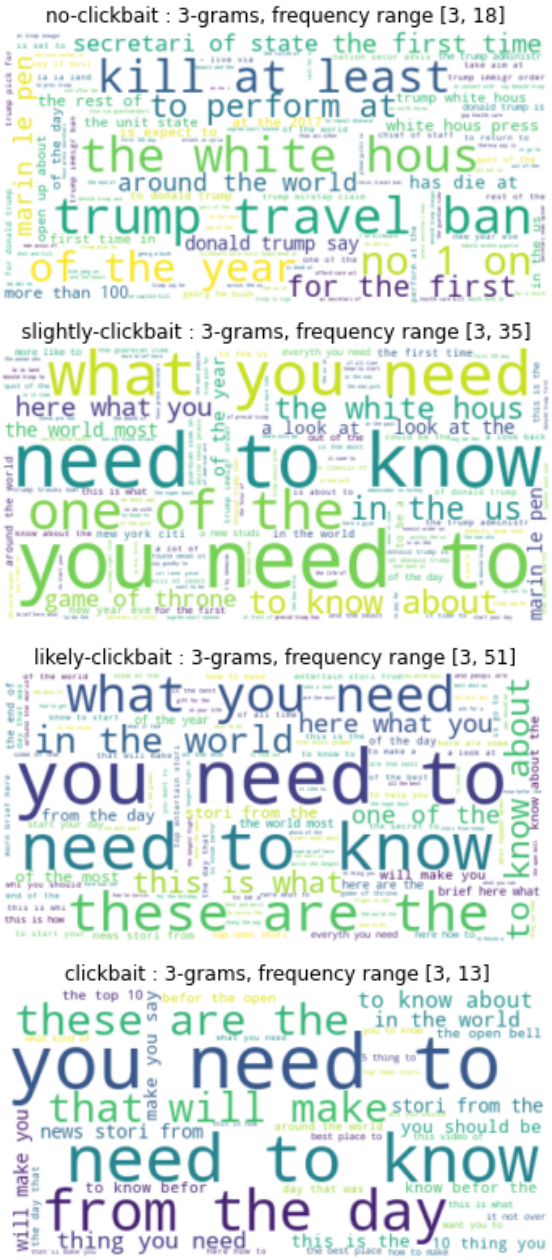

Figure 3: 3-gram wordcloud content of each clickbait scaling grade (0.0 for no-clickbait,0.33, for slightly-clickbait, 0.66 for likely-clickbait,1.0 for clickbait)

4 Methods

In order to adequately steer the text generation towards our attribute a, first we need a good clickbait **classifier** $P(X|a, X')$. For that, encoder based models BERT(Devlin et al., 2018), RoBERTa (Liu

et al., 2019) and MPNet (Song et al., 2020) were taken into consideration.

For steering our clickbait title generation, we have considered and reimplemented two previously existing models:

- **PPLM** (Dathathri et al., 2019) which is using gradient ascent on text generation model $P(X|X')$ to steer model towards classifier's output score $P(a|X, X')$.
- **Fudge** (Yang and Klein, 2021), which is using text generator's $P(X|X')$ decoder's outputs and multiplying them with classifier's output scores $P(a|X, X')$ to steer text generation towards "clickbaitiness" attribute.

In both models, reimplementation included switching **text-generation** from GPT-2 (Radford et al., 2019) model to context-based Pegasus model (Zhang et al., 2019).

Both these models have already done this for generating text based on the attribute classifier, but none of them have ever done generating according to tunable attribute parameter in continuous space, i.e. instead of using classes 0 and 1 for steering, we use the whole space $[0, 1]$.

5 Results

5.1 Clickbait Classifier

As we mentioned above, the state of the art encoder-based models are used for classification based on the text title as the input on the above mentioned Webis dataset (Potthast et al., 2018). Since these models are pre-trained, they already contain some context on how certain words are correlated, but need the fine-tuning for the specific task to give us reasonable results. Fine-tuning is done for 6 epochs and by training it on the binary objective ("clickbait" and "no-clickbait") using binary cross entropy loss. In our case, the best model appeared to be MPNet (Song et al., 2020). Result we get is compared with the Webis 2017 leaderboard in a table 1.

Although our model doesn't achieve the best result in any of the metrics, in total it is one of the top metric results among other classifiers existing on the leaderboard. Therefore, we assume our classifier is good enough to be used for steering text generation in the next step.

5.2 Clickbait text generation

Our text generation works by using large language models in the few-shot setting, fine-tuning our text generator only on an epoch, and conditioning it on a trained classifier. Since our goal is to generate clickbait titles according to the input news article, we are also using a clickbait classifier from the previous step to see how well our title generator works. In the tables 2 and 3, we are showing what titles on news articles containing Kim Jong Un and Elon Musk (separately) look like. We see a couple of examples of how titles look in the original form (in the left column), and when it is being generated by our model (on the right column). Since sometimes our model is not able to generate clickbaity titles as we expect, we are also giving examples where the original titles were clickbait/non-clickbait, but the generated one was non-clickbait according to our clickbait classifier. Besides these two named entities, in the appendix A, we are also providing generated titles on articles containing Barack Obama, Hillary Clinton, combined Macron/Le Pen, Kanye West, the Republican Party, and Democratic Party.

An important part of our work also includes exploring how well our text generation model works for the different clickbait score ranges between $[0, 1]$. To get those title ranges based on the clickbait score, we also had to somehow tune our models to work properly with different hyperparameter

Team name	F1 score	Precision	Recall	Accuracy
Our	**0.678 (2nd)**	**0.707 (3rd)**	**0.65(2nd)**	**0.853 (2nd)**
whitebait	0.565	0.699	0.474	0.826
zingel	0.683*	0.719*	0.65	0.856*
arowana	0.656	0.659	0.654*	0.837
houndshark	0.451	0.478	0.427	0.753
pike	0.604	0.711	0.524	0.836
torpedo	0.65	0.53	0.841	0.785

Table 1: Webis 2017 (Potthast et al., 2018) leaderboard.

Named Entity: Kim Jong Un	
Original non-clickbait	**Generated non-clickbait**
U.S. wants more U.N. sanctions over North Korea's nuclear arms, warns time is short	Tillerson: New economic, diplomatic sanctions 'painful' for North Korea !!!.! America's Diplomacy War
Trump Says 'Major, Major' Conflict With North Korea Possible, But Seeks Diplomacy	'The major conflict' risk with North Korea is 'absolutely possible' Trump says in interview
Original non-clickbait	**Generated clickbait**
Donald Trump praises North Korean leader, saying 'not many 27-year-old men could go in and take over a regime'	Trumpweighs in on whether the world's most crazy leader is rational
Trim Jong-un: North Korean men and women have a choice of just 15 approved hairstyles – but none match Kim Jong-un's distinctive cut	North Korean styles have 15 approved by Communist regime - just 15 male adults?! Damn?! - I tried the 15! - and it didn't fit...
Original clickbait	**Generated non-clickbait**
REVEALED: Kim's plan to kidnap Western tourists from South Korea to hold as hostages if the US attacks the Hermit Kingdom... as an American carrier strike group closes in on the North Korean coastline	North Korean spies trained to kidnap US diplomats and hold them hostage in case of war with the US revealed

Table 2: Mentions of Kim Jong Un in news articles. On the left side are original titles from the news articles, and on the right side are titles generated by our Fudge (Yang and Klein, 2021), and they are labelled as clickbaity depending on our classifier's estimation. More examples in the Appendix.

Named Entity: Elon Musk	
Original non-clickbait	**Generated non-clickbait**
Elon Musk hints at new Tesla feature for Easter eggs	Tesla says it has the most hidden Easter eggs yet it's just getting revealed to the public by Elon Musk
GM's Cruise Poised To Add 1,100 Silicon Valley Autonomous Car Tech Jobs	Could General Motors Remember San Francisco? It's Adding More Than 1,100 Self-Driving Jobs Right Now<n>;]
Original non-clickbait	**Generated clickbait**
Elon Musk Has an Awkward Problem at Tesla: Employee Parking	Elon Musk's 3 Biggest Nightmares About Parking Like Really.
Elon Musk Is About To Launch A Lethal Pathogen Into Space – It Might Just Save Your Life	What could be a killer pathogen in space?, space could! we to the same side and we'll be happy! and more
Original clickbait	**Generated non-clickbait**
These three "killer apps" could help humans colonize Mars	How to Colonize Mars Will Be Based on Three Killer Apps,Experts Say to Mars of
Original clickbait	**Generated clickbait**
10 books that famous CEOs loved reading	21 company CEOs and successful entrepreneurs who read every night: The list books read by the world's top business leaders

Table 3: Mentions of Elon Musk in news articles. On the left side are original titles from the news articles, and on the right side are titles generated by our Fudge (Yang and Klein, 2021), and they are labelled as clickbaity depending on our classifier's estimation. More examples in the Appendix.

settings and correlate them with clickbait scores, for both PPLM (Dathathri et al., 2019) and Fudge

(Yang and Klein, 2021) approach. To explore that, we used **conditional lambda** in Fudge approach (Yang and Klein, 2021), and **stepsize** in PPLM approach (Dathathri et al., 2019), which were also recommended by respective authors to tune for steering towards target attributes. The results are shown in the figure 4 for both approaches. As it can be seen, with increasing hyperparameter intensity, both approaches show increasing clickbait scores overall. In the case of a PPLM approach, (Dathathri et al., 2019), this growth is going with a more fine increase.

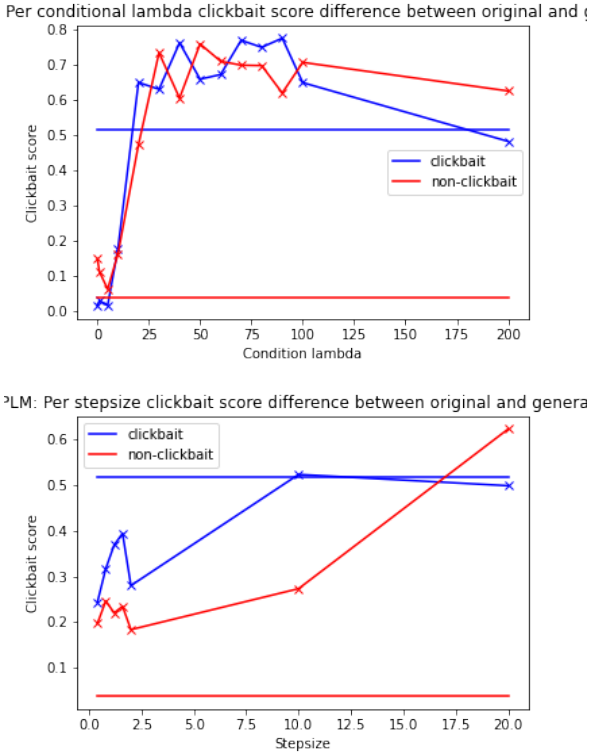

Figure 4: Clickbait score changes with the hyperparameter tuning. Each figure compares generated title scores on a clickbait classifier against the hyperparameter change. The separation between clickbait and non-clickbait titles is shown for the test samples' ground truth clickbait values. There are two horizontal lines on each figure to show the original title score as estimated by our clickbait classifier. For both models, we see the correlation of the relevant hyperparameter with an overall clickbait score grade per subset of a dataset.

On the other hand, while this increase indeed shows continuous growth of clickbaitiness of generated titles, one issue that arises is the consistency of the generated titles with the news articles it is supposed to target. In figure 5, we compare rouge scores of generated titles vs lead-3 baseline (where lead-3 means the first three sentences from each news article). We see that through increased hyperparameter, while clickbaitiness increases, its rouge score also drops. While this could be seen as a good thing, given that the point of a clickbait title is to be far from truthful content in the original article content, in our case it also means that generated titles can look incoherent in both syntax and semantics. This is something that should be a focus of future work and that more fine-tuning is needed for text generation to generate human-readable clickbait titles

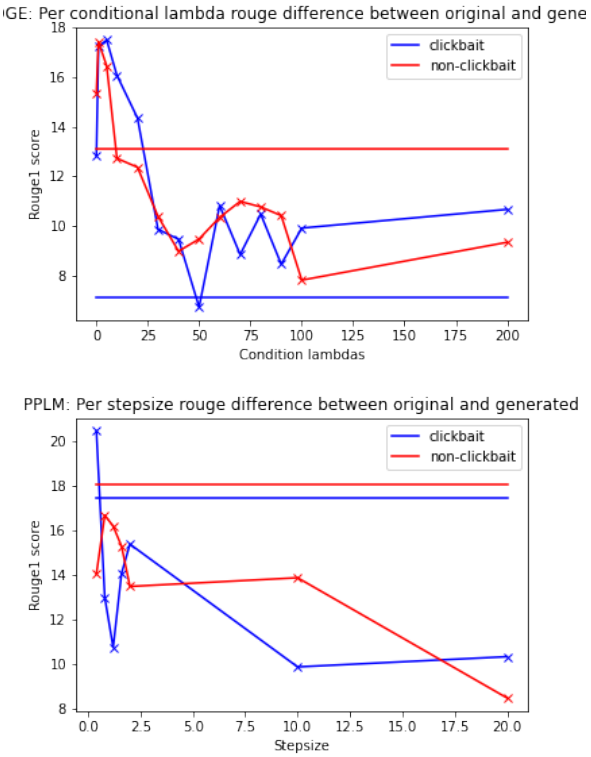

Figure 5: Rouge score change with the hyperparameter tuning. For both models there is a rouge score drop in coherency of the generated titles when compared to the lead-3 baseline. Horizontal lines are rouge scores of ground truth original titles vs lead-3 baseline. Lead-3 means first three sentences of each news article.

6 Conclusion

We have shown how our methods act when tuning hyperparameters, and how it affects final classifier's outputs. There are two major approaches to be used for further improvements. The first one is training classifiers in a different manner compared to our approach. Since we train our classifier on a binary classification task, there is a firm reason to claim that this type of approach is not good enough for fine attribute-tuning to a $[0, 1]$

range as we did, and that it might steer too much towards attribute-preferred tokens while text generating. Another way to consider would be as is done in Fudge (Yang and Klein, 2021), where they learn classifiers on subsequences to predict titles, which in the token-per-token approach could work better towards steering the attribute scores we are looking for (which they also do, but not towards continuous attribute scores, but only binary). The second one would be instead of doing a grid search over relevant hyperparameters in Fudge (Yang and Klein, 2021) and PPLM (Dathathri et al., 2019), we instead used some more sofisticated hyperparameter searches, such as Bayesian optimization or meta-learning algorithms for adequatelly setting up classifiers to give different results for different current title lenghts.

References

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The long-document transformer.

Big Science. 2021. Bigscience: A one-year long research workshop on large multilingual models and datasets.

Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2019. Plug and play language models: A simple approach to controlled text generation.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.

Nitish Shirish Keskar, Bryan McCann, Lav R. Varshney, Caiming Xiong, and Richard Socher. 2019. Ctrl: A conditional transformer language model for controllable generation.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

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*, pages 1498–1507, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: Masked and permuted pre-training for language understanding.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

Kevin Yang and Dan Klein. 2021. FUDGE: Controlled text generation with future discriminators. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics.

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2019. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization.

A Appendix

A.1 N-gram wordcloud

In this part we show most frequent 3-grams found in the dataset. The frequency on non-clickbaity titles is higher because the ratio of non-clickbaity titles is higher as well.

A.1.1 Most frequent 3-grams for no-clickbait

n-gram	frequency
trump travel ban	18
the white hous	17
kill at least	15
no 1 on	12
of the year	12
to perform at	11
marin le pen	9
the first time	9
for the first	9
around the world	9
donald trump say	8
secretari of state	8
the rest of	7
to donald trump	7
more than 100	7
donald trump is	7
the trump administr	7
in the us	7
of the day	7
open up about	7
white hous press	7
trump white hous	7
first time in	7
is expect to	7
has die at	7
the unit state	7
at the 2017	7
trump immigr order	6
rest of the	6
be abl to	6
trump wiretap claim	6
chief of staff	6
live via guardiansport	6
say it will	6
– live via	6
la la land	6
is set to	6
for donald trump	6
quot of the	6
take aim at	6
of the world	6
one of the	6
nation secur advis	6
trump immigr ban	6
trump pick for	6
georg hw bush	6
donald trump inaugur	6
to return to	6
new year eve	6
perform at the	6

A.1.2 Most frequent 3-grams for slightly-clickbait

n-gram	frequency
you need to	35
need to know	31
what you need	21
to know about	20
one of the	20
in the us	18
the white hous	17
here what you	15

n-gram	frequency
game of throne	14
the world most	14
look at the	14
marin le pen	13
a look at	11
in the world	11
new year eve	11
of the year	11
around the world	10
the first time	10
to be a	10
new york citi	10
the trump administr	9
out of the	9
this is what	9
this is the	9
everyth you need	9
of the day	9
for the first	9
more like to	9
a new studi	9
a look back	8
know about the	8
is about to	8
could be the	8
of donald trump	8
a lot of	8
donald trump has	7
the guardian view	7
to the us	7
on new year	7
beauti and the	7
and the beast	7
quot of the	7
guardian view on	7

n-gram	frequency
trump travel ban	7
trump immigr order	7
it time to	7
on donald trump	7
donald trump is	7
suprem court nomine	6

A.1.3 Most frequent 3-grams for likely-clickbait

n-gram	frequency
you need to	51
need to know	38
these are the	25
in the world	21
what you need	21
to know about	18
this is what	16
here what you	15
one of the	15
from the day	14
of the most	13
stori from the	13
will make you	12
here are the	11
brief here what	11
know about the	11
the world most	10
news stori from	10
of the year	10
the end of	9
everyth you need	9
this is how	9
here how to	9
top news stori	9

n-gram	frequency
of the day	9
start your day	9
that will make	7
to know to	7
here are some	7
of all time	7
know to start	7
of the world	7
morn brief here	7
to start your	7
of the best	7
entertain stori from	7

A.1.4 Most frequent 3-grams for clickbait

n-gram	frequency
you need to	13
need to know	10
from the day	7
these are the	7
in the world	6
thing you need	6
to know about	6
that will make	6
will make you	5
stori from the	5
news stori from	5
to know befor	4
around the world	4
the top 10	4
the open bell	4
know befor the	4
befor the open	4
10 thing you	4
the day that	4
you should be	4
this is the	4
make you say	4
day that was	4
here how to	3
top news stori	3
this is how	3
that'll make you	3
this is what	3
what you need	3
best place to	3
the best place	3
how to make	3
5 thing to	3
you to know	3
want you to	3
this video of	3
it not over	3
what kind of	3
whi you should	3

A.2 LDA of topics

We have chosen 18 topics to look at, with coherency score of 0.36.

topic_id	keywords used
0	0.028*"#" + 0.018*"new" + 0.007*"take" + 0.007*"want" + 0.006*"say" + 0.005*"chang" + 0.005*"amp" + 0.005*"video" + 0.005*"song" + 0.004*"street"
1	0.012*"us" + 0.011*"#" + 0.009*"court" + 0.009*"say" + 0.009*"nation" + 0.005*"suprem" + 0.004*"countri" + 0.004*"pull" + 0.004*"russian" + 0.004*"russia"

topic_id	keywords used
2	0.021*"#" + 0.008*"year" + 0.008*"girl" + 0.007*"us" + 0.007*"explain" + 0.006*"say" + 0.005*"new" + 0.005*"women" + 0.004*"surpris" + 0.004*"golden"
3	0.038*"#" + 0.013*"best" + 0.010*"day" + 0.009*"christma" + 0.009*"peopl" + 0.008*"get" + 0.007*"internet" + 0.007*"ever" + 0.006*"year" + 0.006*"top"
4	0.014*"say" + 0.010*"#" + 0.008*"attack" + 0.006*"berlin" + 0.006*"suspect" + 0.006*"kill" + 0.005*"forc" + 0.005*"christma" + 0.005*"via" + 0.005*"new"
5	0.010*"say" + 0.008*"us" + 0.008*"obama" + 0.008*"year" + 0.007*"presid" + 0.007*"#" + 0.006*"new" + 0.006*"could" + 0.006*"attack" + 0.006*"offici"
6	0.017*"know" + 0.014*"here" + 0.012*"get" + 0.010*"say" + 0.009*"need" + 0.007*"#" + 0.006*"call" + 0.006*"us" + 0.005*"much" + 0.005*"could"
7	0.049*"#" + 0.007*"year" + 0.006*"fact" + 0.006*"go" + 0.005*"one" + 0.005*"top" + 0.005*"brand" + 0.005*"new" + 0.004*"first" + 0.004*"import"
8	0.011*"say" + 0.008*"#" + 0.007*"bill" + 0.007*"plan" + 0.006*"wall" + 0.006*"us" + 0.006*"china" + 0.005*"accus" + 0.005*"back" + 0.005*"state"
9	0.032*"#" + 0.008*"presid" + 0.007*"bowl" + 0.007*"super" + 0.007*"year" + 0.005*"us" + 0.005*"icon" + 0.005*"china" + 0.005*"ad" + 0.005*"say"
10	0.011*"#" + 0.008*"via" + 0.007*"time" + 0.007*"night" + 0.006*"still" + 0.006*"world" + 0.006*"around" + 0.006*"work" + 0.005*"new" + 0.005*"show"
11	0.018*"#" + 0.012*"polic" + 0.011*"man" + 0.010*"new" + 0.008*"shoot" + 0.007*"found" + 0.007*"kill" + 0.006*"dead" + 0.005*"near" + 0.005*"face"
12	0.017*"hous" + 0.015*"#" + 0.015*"white" + 0.008*"watch" + 0.007*"via" + 0.006*"come" + 0.005*"press" + 0.005*"live" + 0.005*"famili" + 0.004*"weekend"
13	0.016*"#" + 0.009*"north" + 0.009*"die" + 0.009*"korea" + 0.008*"carri" + 0.006*"star" + 0.006*"rape" + 0.006*"us" + 0.006*"" + 0.005*"south"
14	0.025*"#" + 0.013*"via" + 0.007*"guardiansport" + 0.007*"say" + 0.007*"–" + 0.006*"dog" + 0.006*"game" + 0.006*"save" + 0.005*"new" + 0.005*"could"
15	0.037*"#" + 0.011*"new" + 0.009*"may" + 0.007*"year" + 0.006*"tweet" + 0.006*"presid" + 0.006*"day" + 0.006*"immigr" + 0.006*"ban" + 0.005*"travel"
16	0.060*"#" + 0.009*"say" + 0.008*"million" + 0.006*"presid" + 0.006*"dead" + 0.005*"new" + 0.005*"die" + 0.005*"univers" + 0.004*"report" + 0.004*"school"
17	0.015*"#" + 0.008*"first" + 0.007*"world" + 0.007*"us" + 0.006*"end" + 0.005*"here" + 0.005*"year" + 0.005*"new" + 0.005*"way" + 0.004*"syria"

A.3 Additional generated titles under Named Entity

Below there is a set of tables for various Named Entities mentioned in the news articles. Generated titles, once generated, were a little bit edited by Dusan, to make them more coherent.

Named Entity: Le Pen/Macron	
Original non-clickbait	**Generated non-clickbait**
Macron's rise shows that extremism – whether left or right – is not inevitable	Britain, beware of the impact of a France election disaster' By Andrew Gregory]
Partner of gay policeman murdered in latest Paris terror attack pays tribute as the victim is honoured at ceremony attended by Le Pen and Macron - still bitterly divided on how to deal with jihadist extremists	Paris policeman murdered by ISIS sympathiser remembered at ceremony attended by Macron, Le Pen by all
Original non-clickbait	**Generated clickbait**
Emmanuel Macron Steps Into France's Political Void	This Could Be a Very Unpredictable French Election And The Most Anti-Christian Candidate In Town What?
Original clickbait	**Generated non-clickbait**
What to watch for in the first round of the French presidential election	How Far Is Marine Le Pen Going To Make It To the French Election?!!
France's liberal Macron courts expats in UK in battle for votes	Macron is the French dream come true - at home? – why you need to vote for the one who could turn out....
Original clickbait	**Generated clickbait**
The Pessimist's Guide to 2017	The Pessimist's Guide to 2017 Including Predictions and More:
Inside Le Pen's Growing Army of Discontent	What Liberalism Thinks About Populistismd. p p. the liberalism as a political movement of the way

Table 9: Mentions of French presidential candidates Macron and Le Pen in news articles. On the left side are original titles from the news articles, and on the right side are titles generated by our Fudge (Yang and Klein, 2021), and they are labelled as clickbaity depending on our classifier's estimation. More examples in the Appendix.

Named Entity: Kanye West	
Original non-clickbait	**Generated non-clickbait**
Grammys producer addresses Justin Bieber, Kanye West boycott rumors	Grammys 'Still Talking' With Justin Bieber, Kanye West About Performing 2017 List
Jay Z Salutes Should-Be Rap Honorees in Songwriters Hall of Fame Playlist	Jay Z's Gettin' It On for His Songwriters Hall Of Fame: Listen Playlist Post of 'Otis"
Original non-clickbait	**Generated clickbait**
Celebrities Speaking Out: Are They Helping or Hurting With Anti-Trump Antics?	Are Hollywood Celebrities Hurting Their Own Cause? to the Trump Administration?
2017 Music Preview: 40 Most Anticipated Albums	The Best Albums The Internet Has Already Said For 2017, Official , And More!. to Post With
Original clickbait	**Generated non-clickbait**
Here Are All the 2017 Grammy Award Winners	Find Best Music Video Awards 2017 & Best Pop Solo Performance And Don't Be Late To Watch Me!
Future sounds: the best pop and rock of 2017	BBC 6 Music picks out the 10 best new releases from around the world..... xx
17 Damn Good Albums From 2016	2016's Most Beautiful Songs of the Year, We Belong' More
Original clickbait	**Generated clickbait**
Here's How "Keeping Up With The Kardashians" Dealt With The Paris Robbery	Kim Interview + Life Update + Album Hint: She Felt Such Amputated on This Tour Is That So Real So Kardashian.
T.I. on Kanye West's Trump Meeting: 'What the F*ck You Doing, Man?'	"Don't Just Say Ya'll Might Not Matter": T.I. Explains The Secret Agenda Behind Kanye West, Steve Harvey And More

Table 10: Title of articles mentioning Kanye West. On the left side are original titles from the news articles, and on the right side are titles generated by our Fudge (Yang and Klein, 2021), and they are labelled as clickbaity depending on our classifier's estimation.

Named Entity: Hillary Clinton	
Original non-clickbait	**Generated non-clickbait**
House Oversight Chairman Jason Chaffetz won't seek re-election	US Rep. Jason 'Uncle Todd' Chaffetz, House Committee Heading to 'The World's Biggest Job' Also Won't Run for Reelection Under a Different Party
Trump's Shadow Is Looming Over This Special Election	In First Trump-Tailored Official Test, Dems Are Outraged By The Unpopular President's Turnaround: David Goldman
Original clickbait	**Generated non-clickbait**
EXCLUSIVE – Clinton Sex Accusers Slam 'Hypocritical' Media over Bill O'Reilly Coverage	Victims Shoot Out at Media for Dismissing Bill Clinton's Sexual Assault Charges."/ (Baste). ch.
What to watch for in the first round of the French presidential election	France: What's at Stake in this Year's Biggest Election,
Amber Tamblyn reveals baby name via Hillary Clinton humble-brag	Amber Tamblyn shows off a secret letter from Hillary Clinton her newborn daughter. As one a lot of Hilary Clintons do
Original clickbait	**Generated clickbait**
Hillary Spokesman: Clinton Campaign Didn't Appreciate 'Profound' Impact of the 'Breitbart Effect'	Brevity of the 2016 Will Once You're Out of It.. [sic]] Here's Why A Key Political Advisor Who Just Met the Trump Campaign Was Dismayed That
Trump's piano guy and cello player just trolled Hillary fans, bigly	Watch The Piano Guys get Donald Trump's team in a mess by playing a nasty and anti-climax song The minute they'll be here
A Troll Outside Trump Tower Is Helping To Pick Your Next Government	Donald Trump's Internet Troll Helping Candidate, Sources a - To -

Table 11: Title of articles mentioning Hillary Clinton. On the left side are original titles from the news articles, and on the right side are titles generated by our Fudge (Yang and Klein, 2021), and they are labelled as clickbaity depending on our classifier's estimation.

Named Entity: Barack Obama	
Original non-clickbait	**Generated non-clickbait**
Fabio: California Is a 'Mess' Because of Liberal Policies Insider	'Tucker Carlson Tonight': I'm in California and I'm Fine, California Is 'a Mess'!"
Obama's $400G Wall Street speech leaves liberal base stunned	Left's anger over Wall Street speech from then-President Obama
Surgeon General removed from post by Trump administration, replaced by deputy	Surgeon General, who Appointed Obama; Says He's Stepping Down
'They Almost Won Something': Gutfeld Jabs Dems for Celebrating GA Election Loss Insider	Fox News Host Loves to Make Fun of Liberals Celebrating After Republican Winner 'Almost Wins' Special Election. Subscribe
Original non-clickbait	**Generated clickbait**
Special Immigrant Visa Worth Billions Faces D-Day	Getting an EB-5 Visa Is Very Hard, but What's The Deal? with
Trump's Trainwreck	Trump's Republicans were so desperate to undo Obama they saw something they weren't expecting and they died from it .
Original clickbait	**Generated non-clickbait**
A young prince is reimagining Saudi Arabia. Can he make his vision come true?	Future of Saudi Arabia is look inward, not outward
Donald Trump Thinks Susan Rice Committed a Crime	Trump: I Think Susan Rice Did Criminal Activity Biggest Story of Our Time ?

Table 12: Title of articles mentioning Barack Obama. On the left side are original titles from the news articles, and on the right side are titles generated by our Fudge (Yang and Klein, 2021), and they are labelled as clickbaity depending on our classifier's estimation.

Named Entity: Democratic Party	
Original non-clickbait	**Generated non-clickbait**
U.S. wants more U.N. sanctions over North Korea's nuclear arms, warns time is short	New sanctions on North Korea must come now Tillerson warns at Security Council !
'I know he is alive': wife of Taiwan activist seized by China pleads for release	Women lobbying Europe and the US for release of Chinese activist wife man detained her... a special for our freedom we
Original non-clickbait	**Generated clickbait**
Mother and daughter die after getting stuck in a sauna	Mom and daughter die in freak accident in their own saunas This one:
Trump's Trainwreck	The Trumpcare Myth Is More Rotten Than You Think (Part 1
Ivanka Trump is a reassuring presence in an unpredictable White House	Why Ivanka has a West Wing office anything else?-, then- two."-.. ()-
Key moments from today's hearing with FBI Director Comey and NSA Director Rogers	Comey, Rogers testify in House hearing: Top moments) It's clear the big story of the hearing was indeed the continuing Russia probe
Original clickbait	**Generated non-clickbait**
Trumplomacy: Bromance with an autocrat	The Trump administration is to welcome Egypt's brutal autocrat back to Washington
Ex-Trump University student wants the president's apology	Donald Trump's $25-million settlement over fraud allegations may be put on hold judge decides
Original clickbait	**Generated clickbait**
Here's the most important exported good from every state	How big is the global economy?
The Daily 202: The price of loyalty to Trump is high and rising	In And So It Goes: The Secret Behind Donald Trump's Heightened self-Control Says My Times And You Don?

Table 13: Title of articles mentioning Democratic Party. On the left side are original titles from the news articles, and on the right side are titles generated by our Fudge (Yang and Klein, 2021), and they are labelled as clickbaity depending on our classifier's estimation.

Named Entity: Republican Party	
Original non-clickbait	**Generated non-clickbait**
Trump eyes bid to split up 9th Circuit court amid sanctuary city, travel ban disputes	Trump says he considers 9th Circuit split takes aim at opponents at the ; and the will of the people
Democrats open to more border security money to avoid funding standoff	Trump may be trying to back off from border wall funding threat Dems shine light on compromise too corner Subscribe now or save now
Home for recess, lawmakers face fiery town hall crowds on Trump, abortion and more	Town Hall fury is bacl, get ready for another week of anger @CBSN at this time
Trump Makes His First Big Changes To Obamacare	Trump Administration Wants to Make 'Obamacare' Exchanges More Affordable, Less Toxic But Untrue Insurance Companies Are Worried
Original non-clickbait	**Generated clickbait**
By the numbers: Why millions go uninsured under Republicans' Obamacare alternative	What would happen to Obamacare under the GOP bill? here are the numbers to
Trump 'Historic' Defense Increase Is Barely Above What Obama Planned – And Faces Similar Obstacles	The Pentagon's New Plan Isn't All Big, It's Just Slightly Oversized]
Former WWE wrestler stops armed robber in Florida	Former WWE Wrestler Shamed, Captured An Armed Robbery Puzzlingly... Continue and more...
Original clickbait	**Generated non-clickbait**
RNC Chair: GOP Base Will 'Walk Away' in Midterms If Border Wall Is Not Built	RNC Chair: If Congress Doesn't Help Trump Build Border Wall, Voters Will Walk Away"': Ingraham
Los Angeles tops list of dog attacks on mail carriers	America's Top Dog Attacks: Los Angeles, Philadelphia, San Francisco
Ex-Trump University student wants the president's apology	Trump may settle class-action lawsuit over Trump University lawsuit, Sherri Simpson back — More Read] —
Trump calls out Freedom Caucus members in tweetstorm	Trump goes to war with the House Freedom Caucus again, Ryan says more GOP's problem
Original clickbait	**Generated clickbait**
The Border Adjustment Tax For Dummies: Who Will Pay For The Wall?	What a Really Strange Tax System It Is Mexico... that ...
Winners and losers of the biggest state political battles of 2016	Nine of 2016's Biggest State Political Battles

Table 14: Title of articles mentioning Republican Party. On the left side are original titles from the news articles, and on the right side are titles generated by our Fudge (Yang and Klein, 2021), and they are labelled as clickbaity depending on our classifier's estimation.